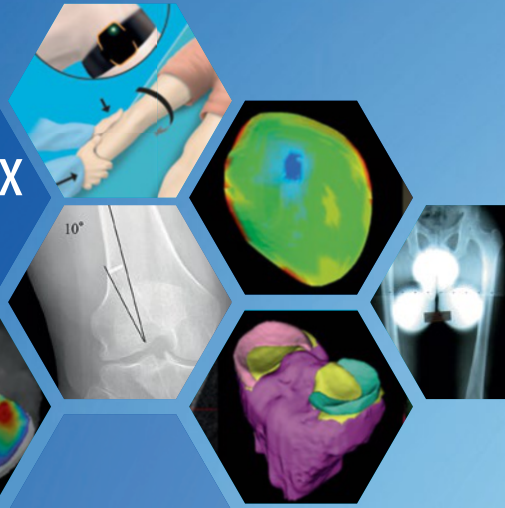


Volker Musahl · Jón Karlsson
Michael T. Hirschmann
Olufemi R. Ayeni · Robert G. Marx
Jason L. Koh
Norimasa Nakamura
Editors



Basic Methods Handbook for Clinical Orthopaedic Research

A Practical Guide and Case Based
Research Approach



 Springer

Basic Methods Handbook for Clinical Orthopaedic Research

Volker Musahl • Jón Karlsson
Michael T. Hirschmann
Olufemi R. Ayeni • Robert G. Marx
Jason L. Koh • Norimasa Nakamura
Editors

Basic Methods Handbook for Clinical Orthopaedic Research

A Practical Guide and Case Based
Research Approach



Editors

Volker Musahl
UPMC Rooney Sports Complex
University of Pittsburgh
Pittsburgh, PA
USA

Michael T. Hirschmann
Department of Orthopaedic Surgery
and Traumatology
Kantonsspital Baselland
(Bruderholz, Laufen und Liestal)
Bruderholz
Switzerland

Robert G. Marx
Hospital for Special Surgery
New York, NY
USA

Norimasa Nakamura
Institute for Medical Science in Sports
Osaka Health Science University
Osaka
Japan

Jón Karlsson
Department of Orthopaedics
Sahlgrenska Academy
Gothenburg University
Sahlgrenska University Hospital
Gothenburg
Sweden

Olufemi R. Ayeni
McMaster University
Hamilton, ON
Canada

Jason L. Koh
Department of Orthopaedic Surgery
NorthShore University HealthSystem
Evanston, IL
USA

ISBN 978-3-662-58253-4 ISBN 978-3-662-58254-1 (eBook)
<https://doi.org/10.1007/978-3-662-58254-1>

Library of Congress Control Number: 2018966868

© ISAKOS 2019

This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors, and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer-Verlag GmbH, DE part of Springer Nature.

The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Contents

Part I Evidence Based Medicine in Orthopaedics

- 1 What Is Evidence-Based Medicine? 3**
Eleonor Svantesson, Eric Hamrin Senorski, Jón Karlsson,
Olufemi R. Ayeni, and Kristian Samuelsson
- 2 What Is the Hierarchy of Clinical Evidence? 11**
Vishal S. Desai, Christopher L. Camp, and Aaron J. Krych
- 3 Bias and Confounding 23**
Naomi Roselaar, Magaly Iñiguez Cuadra, and Stephen Lyman
- 4 Ethical Consideration in Orthopedic Research 33**
Jason L. Koh and Diego Villacis
- 5 Conflict of Interest 43**
Michael Hantes and Apostolos Fyllos
- 6 Ethics in Clinical Research 49**
Naomi Roselaar, Niv Marom, and Robert G. Marx

Part II How to Get Started with Clinical Research?

- 7 How to Get Started: From Idea to Research Question 57**
Lachlan M. Batty, Timothy Lording, and Eugene T. Ek
- 8 How to Write a Study Protocol 65**
Lukas B. Moser and Michael T. Hirschmann
- 9 The Ethical Approval Process 75**
Karren Takamura and Frank Petrigliano
- 10 How to Assess Patient's Outcome? 83**
Yuichi Hoshino and Alfonso Barnechea
- 11 Basics of Outcome Assessment in Clinical Research 89**
Monique C. Chambers, Sarah M. Tepe,
Lorraine A. T. Boakye, and MaCalus V. Hogan
- 12 Types of Scoring Instruments Available 97**
José F. Vega and Kurt P. Spindler

13	Health Measurement Development and Interpretation	111
	Andrew Firth, Dianne Bryant, Jacques Menetrey, and Alan Getgood	
14	How to Document a Clinical Study and Avoid Common Mistakes in Study Conduct?	121
	Caroline Mouton, Laura De Girolamo, Daniel Theisen, and Romain Seil	
15	Framework for Selecting Clinical Outcomes for Clinical Trials	133
	Adam J. Popchak, Andrew D. Lynch, and James J. Irrgang	
16	Advances in Measuring Patient-Reported Outcomes: Use of Item Response Theory and Computer Adaptive Tests . .	143
	Andrew D. Lynch, Adam J. Popchak, and James J. Irrgang	

Part III Basics in Statistics: Statistics Made Simple!

17	Common Statistical Tests	153
	Stephan Bodkin, Joe Hart, and Brian C. Werner	
18	The Nature of Data	163
	Clair Smith	
19	Does No Difference Really Mean No Difference?	171
	Carola F. van Eck, Marcio Bottene Villa Albers, Andrew J. Sheean, and Freddie H. Fu	
20	Power and Sample Size	185
	Stephen Lyman	
21	Visualizing Data	193
	Stephen Lyman, Naomi Roselaar, and Chisa Hidaka	

Part IV Basic Toolbox for the Young Clinical Researcher

22	How to Prepare an Abstract	209
	Elmar Herbst, Brian Forsythe, Avinesh Agarwalla, and Sebastian Kopf	
23	How to Make a Good Poster Presentation	219
	Baris Kocaoglu, Paulo Henrique Araujo, and Carola Francisca van Eck	
24	How to Prepare a Paper Presentation?	227
	Timothy Lording and Jacques Menetrey	
25	How to Write a Clinical Paper	235
	Brendan Coleman	
26	How to Write a Book Chapter	243
	Thomas R. Pfeiffer and Daniel Guenther	

27 How to Write a Winning Clinical Research Proposal? 249
 Christian Lattermann and Janey D. Whalen

28 How to Review a Clinical Research Paper? 255
 Neel K. Patel, Marco Yeung, Kanto Nagai, and Volker Musahl

Part V How to Perform a Clinical Study: A Case Based Approach

**29 Level 1 Evidence: A Prospective Randomized
 Controlled Study 265**
 Seper Ekhtiari, Raman Mundi, Vickas Khanna,
 and Mohit Bhandari

30 Level 1 Evidence: Long-Term Clinical Results 285
 Daisuke Araki and Ryosuke Kuroda

31 Level 2 Evidence: Prospective Cohort Study 289
 Naomi Roselaar, Niv Marom, and Robert G. Marx

32 Level III Evidence: A Case-Control Study 295
 Andrew D. Lynch, Adam J. Popchak, and James J. Irrgang

33 Level 4 Evidence: Clinical Case Series 301
 Mitchell I. Kennedy and Robert F. LaPrade

**34 How to Perform a Clinical Study: Level 4
 Evidence—Case Report 307**
 Andrew J. Sheean, Gregory V. Gasbarro,
 Nasef M. N. Abedelatif, and Volker Musahl

35 Level 5: Evidence 313
 Seán Mc Auliffe and Pieter D’Hooghe

Part VI How to Perform a Review Article?

36 Type of Review and How to Get Started 323
 Matthew Skelly, Andrew Duong, Nicole Simunovic, and
 Olufemi R. Ayeni

Part VII How to Perform a Systematic Review or Meta-analysis?

**37 What Is the Difference Between a Systematic Review
 and a Meta-analysis? 331**
 Shakib Akhter, Thierry Pauyo, and Moin Khan

38 Reliability Studies and Surveys 343
 Kelsey L. Wise, Brandon J. Kelly, Michael L. Knudsen, and
 Jeffrey A. Macalena

39 Registries 359
 R. Kyle Martin, Andreas Persson, Håvard Visnes,
 and Lars Engebretsen

Part VIII How to Perform an Economic Health Care Study?

- 40 How to Perform an Economic Healthcare Study** 373
Jonathan Edgington, Xander Kerman, Lewis Shi, and
Jason L. Koh

Part IX Multi-Center Study: How to Pull It Off?

- 41 Conducting Multicenter Cohort Studies:
Lessons from MOON** 383
José F. Vega and Kurt P. Spindler
- 42 MARS: The Why and How of It** 391
Rick W. Wright, Amanda K. Haas, and Laura J. Huston
- 43 Multicenter Study: How to Pull It Off? The PIVOT Trial** 403
Eleonor Svantesson, Eric Hamrin Senorski, Alicia Oostdyk,
Yuichi Hoshino, Kristian Samuelsson, and Volker Musahl
- 44 Conducting a Multicenter Trial: Learning from the
JUPITER (Justifying Patellar Instability Treatment
by Early Results) Experience** 415
Jason L. Koh, Shital Parikh, Beth Shubin Stein,
and The JUPITER Group
- 45 How to Organise an International Register in Compliance
with the European GDPR: Walking in the Footsteps
of the PAMI Project (Paediatric ACL Monitoring Initiative)** . . . 427
Daniel Theisen, Håvard Moksnes, Cyrille Hardy,
Lars Engebretsen, and Romain Seil

Part X Helpful Further Information

- 46 Common Scales and Checklists in Sports Medicine Research** . . . 437
Alberto Grassi, Luca Macchiarola, Marco Casali,
Ilaria Cucurnia, and Stefano Zaffagnini
- 47 A Practical Guide to Writing (and Understanding)
a Scientific Paper: Meta-Analyses** 471
Alberto Grassi, Riccardo Compagnoni, Kristian Samuelsson,
Pietro Randelli, Corrado Bait, and Stefano Zaffagnini
- 48 A Practical Guide to Writing (and Understanding)
a Scientific Paper: Clinical Studies** 499
Riccardo Compagnoni, Alberto Grassi, Stefano Zaffagnini,
Corrado Bait, Kristian Samuelsson, Alessandra Menon, and
Pietro Randelli
- 49 Reporting Complications in Orthopaedic Trials** 507
S. Goldhahn, Norimasa Nakamura, and J. Goldhahn

50	Understanding and Addressing Regulatory Concerns in Research	515
	Jason L. Koh, Denise Gottfried, Daniel R. Lee, and Sandra Navarrete	
51	What Is Needed to Make Collaboration Work?	533
	Richard E. Debski and Gerald A. Ferrer	
52	A Clinical Practice Guideline	537
	Aleksei Dingel, Jayson Murray, James Carey, Deborah Cummins, and Kevin Shea	
53	How to Navigate a Scientific Meeting and Make It Worthwhile? A Guide for Young Orthopedic Surgeons	551
	Darren de SA, Jayson Lian, Conor I. Murphy, Ravi Vaswani, and Volker Musahl	
54	How to Write a Scientific Article	561
	Lukas B. Moser and Michael T. Hirschmann	
55	Common Mistakes in Manuscript Writing and How to Avoid Them	579
	Eleonor Svantesson, Eric Hamrin Senorski, Kristian Samuelsson, and Jón Karlsson	

About the Editors

Editors



Volker Musahl, MD is Associate Professor and Chief of Sports Medicine in the Department of Orthopaedic Surgery at the University of Pittsburgh and the Deputy Editor-in-chief of KSSTA. He is one of the top 10 published ACL scientists worldwide and co-Principal investigator for the multi-center STaR trial (Surgical Timing and Rehabilitation for multiple knee ligament injury) and the POETT study (Exercise Therapy for rotator cuff tears). He is the co-head team physician for the University of Pittsburgh Football and Fellowship Director for the Sports Medicine and Shoulder Fellowship at the University of Pittsburgh.



Jón Karlsson, MD, PhD is Professor of Orthopaedic Surgery at the University of Gothenburg, Sweden, and Senior Consultant at the Sahlgrenska University Hospital. He has published more than 400 scientific papers, more than 100 book chapters and is the author/co-editor of 40 books on Orthopaedics and Sports Traumatology. He has mentored more than 50 PhD students in their scientific work. He is currently Secretary of ISAKOS Board of Directors and former chair of ISAKOS Scientific Committee. He is currently Editor-in-Chief of KSSTA (Knee Surgery Sports Traumatology Arthroscopy), one of the leading journals in the category of knee surgery and sports trauma. He has published

several papers on research methodology and evidence-based medicine. He has been team physician for IFK Göteborg, soccer club for more than 30 years.



Michael T. Hirschmann, MD is Professor of Orthopaedic Surgery and Traumatology at the University of Basel, Switzerland. At Kantonsspital Baselland (Bruderholz, Liestal, Laufen) he is Co-Chair of Orthopaedic Surgery and Traumatology, Head of knee surgery, and DKF Head of knee research. He has published over 250 academic peer-reviewed articles and book chapters. He was also editor of numerous books dealing with knee surgery. During the last years, he and his team have won more than 30 research awards for their contributions to knee surgery. His main clinical and research interest is all about knee surgery, in particular sports injuries and the treatment of the degenerative knee including osteotomy and primary and revision knee arthroplasty. In addition, he serves as Deputy Editor-in-Chief of the Knee Surgery Sports Traumatology Arthroscopy Journal (KSSTA) for which he also annually runs a Basic Science Writing Course. He is also on the editorial board of numerous other journals.



Olufemi R. Ayeni, MD, PhD, FRCSC is an Associate Professor in the Division of Orthopaedic Surgery at McMaster University. He has published over 230 academic peer-reviewed articles, numerous book chapters, and secured over one million dollars in research funding. He has presented internationally on sports medicine-related topics and evidence-based medicine. He is currently the Medical Director of the Hamilton Tiger Cats of the Canadian Football League and Fellowship Director for the Sports Medicine and Arthroscopy Fellowship at McMaster University.



Robert G. Marx, MD, MSc is Professor of Orthopedic Surgery and vice-chair of Orthopedic Surgery at Hospital for Special Surgery/ Weill Cornell Medical College in New York City. He has published four books and over 200 peer-reviewed scientific articles. Dr. Marx is Deputy Editor for sports medicine and Associate Editor for Evidence-Based Orthopedics for the Journal of Bone and Joint Surgery. He has previously served as chairman of the scientific committee and also on the Board of Directors of ISAKOS.

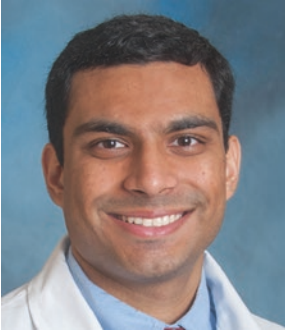


Jason L. Koh, MD, MBA is Chairman of the Department of Orthopaedic Surgery and Director of the Orthopaedic Institute at NorthShore University HealthSystem, Chicago. He is also Clinical Professor of Surgery at the University of Chicago Pritzker School of Medicine and Adjunct Professor at the Northwestern University McCormick School of Engineering. His research regarding tissue engineering, cartilage, rotator cuff, and ligament has been recognized with numerous honors and awards. Dr. Koh has held leadership positions with multiple orthopaedic societies and has published over 100 papers and book chapters. He has served as team physician for the Chicago Cubs, Chicago Fire Soccer, and the Joffrey Ballet.



Norimasa Nakamura, MD, PhD is an orthopaedic surgeon and Professor at the Institute for Medical Science in Sports, Osaka Health Science University, Osaka City, Japan. Dr. Nakamura is a Fellow of the Royal College of Surgeons (FRCS) (England). He is President of the International Cartilage Repair Society (ICRS) and a former chair of the Scientific Committee of ISAKOS. He is a member of the editorial boards of various leading journals in the field of orthopaedics and author of more than 120 peer-reviewed papers.

Associate Editors



Neel K. Patel, MD is an orthopaedic surgery resident in the Clinician-Scientist track at the University of Pittsburgh Medical Center. He completed his undergraduate education at the Massachusetts Institute of Technology and his medical education at the Geisel School of Medicine at Dartmouth. His research interests include shoulder and ankle biomechanics and complex multi-ligament knee injuries.



Darren de SA, MD, FRCSC is a Canadian orthopaedic surgeon from Hamilton, Ontario, Canada, whose clinical scope includes adult and paediatric sports medicine, arthroscopic surgery, and trauma. He has a special interest in complex knee reconstruction and shoulder surgery. Darren completed both his medical and orthopaedic training at McMaster University, and subsequently completed Fellowships in Sports Medicine/Arthroscopy and Adolescent Trauma at the University of Pittsburgh and Western University, respectively. Darren has contributed to a productive research and medical education program focused on outcomes in athletic injuries and non-arthritic conditions of the knee, shoulder, and hip. He is an Editorial Board Member for Arthroscopy and is a manuscript reviewer for both Arthroscopy and KSSTA. His contributions have been recognized at the institutional and national level.

Part I

Evidence Based Medicine in Orthopaedics



What Is Evidence-Based Medicine?

1

Eleonor Svantesson, Eric Hamrin Senorski,
Jón Karlsson, Olufemi R. Ayeni,
and Kristian Samuelsson

1.1 History of EBM

Throughout the history of medicine, it is obvious that certain treatment strategies have been adapted based on preferences, belief, and rationalism. Some of this “knowledge” was conveyed by senior clinicians to the next generations of clinicians and thereby remained as the way to practice medicine without questioning the performance of it. Gradually, some pioneers started to alter this perspective by advocating that empirical evidence should be incorporated in the practice of medicine, indicating that a scientific approach, including observation and critical

appraisal of the current evidence, was fundamental for a trustworthy conclusion on the optimal care of patients. One of the most famous early clinical trials was James Lind’s study of scurvy in the British navy, which was published in 1753 [16]. Despite this early recorded “trial”, it took as long as until 1962 for the US Food and Drug Administration to declare the Kefauver-Harris Act which declared that it was legally required to perform clinical trials involving human beings prior to establishing claims regarding drug efficacy [18]. As clinicians and researchers of today, it appears obvious that an establishment of such claims entails years of investigation. Any scenario other than performing thorough preclinical and clinical studies before this would be unthinkable for us. We should therefore be reminded to reflect over the relatively fast development in the empirical assessment of evidence that our field has undergone. The question arises, what is the reason for this development? What made us abandon the collection of uncontrolled experiences as the major foundation of clinical decision-making and, instead, strive toward the practice of evidence-based medicine (EBM)? Or has it really happened?

In the beginning of the twentieth century, the orthopedic surgeon Earnest A. Codman at the Massachusetts General Hospital in Boston proposed the end result system [6], which could be considered as a groundbreaking system toward modern EBM. Dr. Codman meant that only by evaluating the outcome, “the end result” of a

E. Svantesson (✉)

Department of Orthopaedics, Institute of Clinical Sciences, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

E. H. Senorski

Department of Health and Rehabilitation, Institute of Neuroscience and Physiology, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

J. Karlsson · K. Samuelsson

Department of Orthopaedics, Institute of Clinical Sciences, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

Department of Orthopaedics, Sahlgrenska University Hospital, Mölndal, Sweden
e-mail: jon.karlsson@telia.com

O. R. Ayeni

Division of Orthopaedic Surgery, McMaster University, Hamilton, ON, Canada
e-mail: ayenif@mcmaster.ca

treatment, it was possible to assess the clinical effectiveness of such a treatment. He advocated that it was required to monitor every patient who received treatment for a follow-up time long enough to establish if the treatment led to a satisfactory outcome. If a treatment failure was determined based on the records of outcome, he meant that the reason for the failure should be investigated and proper actions should be undertaken to prevent similar future failures. This was a radical idea in the early twentieth century, and Dr. Codman's work was encountered by such a strong criticism that he lost his staff position at the Massachusetts General Hospital. Opposingly, we nowadays consider Dr. Codman's approach a milestone for medicine based on empirical evidence. The term EBM, which has been regarded as one of the most important paradigm shifts in medical history [27], was defined in 1991 by Gordon Guyatt as a component of the medical residency program at McMaster University in Hamilton, Ontario, Canada [4]. The concept aimed to educate clinicians on how to perform and interpret scientific evidence in terms of assessment of credibility, critical appraisal of the results, and integration of scientific evidence in the everyday work [10]. This concept was subsequently adapted worldwide and several guides, functioning as a foundation of the understanding of the EBM concept, have been published [13, 15, 26]. One contributing factor for the almost explosive entrance and integration of the EBM was the rapid evolvement of modern technology during this time, which enabled the field of informatics to grow tremendously with, for example, large online databases and scientific journals. Therefore, once the concept was established, the requisites of implementing EMB in a variety of specialties enabled the emergence of a new era of a scientific approach to the practice of medicine.

1.2 What Defines the Practice of EBM?

The practice of EBM entails the integration of the current best evidence in the clinical decision-making process in terms of the care of every indi-

vidual patient. Thus, it is about using scientific evidence conscientiously and objectively for this purpose. To ascertain this, it is required that a sufficient amount of clinically relevant sources of information is acquired, which may include published literature such as basic science research, clinical trials, diagnostic testing, predictive factors, the efficacy of therapeutic interventions, etc. However, it could also include individual observations and expert opinions. The expertise of a highly experienced clinician should also be incorporated into the concept of EBM. EBM is about using the accumulated scientific and statistical knowledge derived by several of the aforementioned sources of information. Moreover, it involves using critical appraisal of such sources to evaluate: *what does the strongest evidence suggest in terms of decision-making in this clinical situation?* Nonetheless, to solely rely on scientific evidence without clinical expertise for clinical decision-making would be a misconception of the EBM concept. An important component is also to incorporate patients' perspectives and values, which is in line with providing the best care for every *individual* patient. Establishing a good relationship between the medical professional and the patient is important for this purpose, and the practice of EBM should be characterized by the synergistic values between the brain and the heart [21]. Thus, EBM is characterized by three equally important fundamental principles: the best available research, the clinical experience, and the patient's perspective [24].

1.3 The Best Available Evidence

The large amount of available literature needs a systematic approach to evaluate and synthesize data. Before the term EBM was even established, efforts to describe how to systematically examine the scientific literature for "critical appraisal" and for extraction of evidence were made by David Sackett of McMaster University [24]. The idea of summarizing evidence was developed as a cornerstone of EBM, and creating such summaries has been facilitated by the rapid development of venues for finding information, as well as

advanced degrees in librarian science to help extract data efficiently. Today, finding the best possible evidence is closer than ever with various software programs and databases that can generate information almost instantly. There are several valuable sources that provide clinicians with the best available evidence, including systematic reviews and evidence-based clinical guidelines. Perhaps the most extensive database providing in-depth systematic reviews on a numerous of topics is the Cochrane Database, which also comprises a list of randomized clinical trials in orthopedics and many other subspecialty areas.

1.3.1 The Hierarchy of Evidence

The hierarchy of evidence needs to be appreciated to facilitate interpretation of the large amount of literature on a topic. Moreover, a systematic quality appraisal needs to be carried out when searching for the best possible evidence because not all research is reliable research. The hierarchy of evidence has been established as levels of evidence, which mainly depends on the quality of study design and the expected risk of bias [5, 23, 28]. Although there are several versions of the hierarchy of evidence, the most frequently used version is the one available from the Oxford Centre for Evidence-Based Medicine website, www.cebm.net. In this version, the randomized controlled trials (RCTs) are regarded as the highest level of evidence (level 1) of individual studies, while expert opinions and uncontrolled studies are considered as of the lowest level (level 5). Controlled observational studies intake a position in the middle of these two, and in turn, the levels are further depending on whether a prospective or retrospective approach to the study design has been undertaken. Moreover, the assessment of level of evidence differs slightly depending on the study type, and specific criteria have been established for each study type. According to the systemic approach of the level of evidence assessment proposed in the version by the Oxford Centre for Evidence-Based Medicine, study types are stratified in the following groups: therapeutic studies, prognostic

studies, diagnostic studies, prevalence studies, and economic/decision analyses. Thus, each study type has its own system for determining the level of evidence, and subgroups within a certain level have also been established (e.g., level 1a, 1b, etc.). Fact Box 1.1 summarizes the hierarchy of evidence for therapeutic studies.

Fact Box 1.1: Levels of Evidence

1	Meta-analyses Systematic reviews Randomized controlled trials
2	Cohort studies
3	Case-control studies
4	Retrospective case series
5	Expert opinions

The hierarchy of evidence is proposed to reflect the applicability, reproducibility, and generalizability of a study [9], and a study of high level of evidence should therefore show superiority in terms of these factors. Grant et al. [11] conducted a systematic review to evaluate the level of evidence among published articles in three major journals of sports medicine over the past 15 years. It was concluded that the percentage of level 1 and 2 studies had increased over time, and in 2010 nearly 25% of all studies were level 1 or 2. Although level 4 and 5 studies had decreased over time, these were still the most common level of evidence among sports medicine literature (53% in 2010) [11]. Similarly, another systematic review, specifically investigating the level of evidence among literature related to anterior cruciate ligament reconstruction, concluded that a minority of published literature between 1995 and 2011 were of level 1 evidence (approximately 10%) [25]. However, it is important to point out that EBM does not solely rely on level 1 RCTs, which is a common misunderstanding of the EBM [3]. All types of study designs contribute to EBM, where the strength and disadvantages of every study design need to be considered when accumulating evidence. A valuable instrument for rating evidence quality is the Grading of Recommendations

Assessment, Development, and Evaluation (GRADE) system [2]. The GRADE system has changed the way to address the credibility of various aspects in studies and has provided a standardized fashion to evaluate current evidence. The GRADE system includes not only appraisal of the study design but also a stringent evaluation of risk of bias, precision, variability in results between studies, applicability, effect size, and dose-response gradients. The GRADE rating system is presented in Fact Box 1.2. When systematically synthesizing the results of studies, the GRADE system ensures that an in-depth assessment is undertaken and enables information from different types of study designs to be evaluated, regardless of each study's level of evidence. This is important since it precludes a myopic focus on the confidence in results based solely on study design, for example, the RCTs. The RCTs remain the gold standard for determining the efficacy of an intervention; however, the RCTs are not without limitations. For example, the generalizability of the results from RCTs must be critically evaluated since patient enrollment is based on strict criteria and the trials are often performed in one part of the world, perhaps in highly specialized centers. For evaluating the effectiveness of an intervention in a "real-world population," studies of observational design are instead an important asset. Thus, we should acknowledge the strengths and limitations of each study design, and evidence should be established based on the cumulative results from several types of study designs.

Fact Box 1.2: The GRADE Rating System

Code	Quality of evidence	Definition
A	High	It is unlikely that further research will change current confidence in the estimate of effect. There are consistent results in: <ul style="list-style-type: none"> • Several studies of high quality • One large high-quality multicenter trial

Code	Quality of evidence	Definition
B	Moderate	Further research is likely to impact or change current confidence in the estimate of effect based on: <ul style="list-style-type: none"> • One high-quality study • Several studies with some limitations
C	Low	It is very likely that further research will impact or change current confidence in the estimate of effect based on: <ul style="list-style-type: none"> • One or more studies with severe limitations
D	Very low	The estimate of effect is very uncertain based on: <ul style="list-style-type: none"> • Expert opinion • No direct research evidence • One or more studies with very severe limitations

1.3.2 Evaluation of Hypothesis Testing

The evaluation of study power is an important part of the EBM concept. An underpowered trial is subject to beta errors (Type II errors), and it may have detrimental effects if such trials are to impact clinical practice. A Type II error is present when a study concludes no difference between two interventions when such a difference in fact exists. When conducting a study, the investigators should aim to have a sample size large enough to minimize the risk of Type II errors while maximizing the probability to conclude a difference between two interventions when a real difference exists. This is commonly referred to as the power of a study. Investigators generally accept a probability of 20% to conclude Type II errors, meaning that the correct conclusion will be present in 80% of the time, corresponding to a study power of 80% ($1 - \beta$). A power analysis is performed prior to study start to ensure that a sample size large enough is enrolled in the trial. A power analysis also ensures that the study is feasible, since too large of a sample size can lead

to wasted resources, wasted time, and incomplete studies. Furthermore, enrollment of more patients than necessary in a trial would be ethically incorrect. A priori power analysis is the most valid method; however, such an analysis could also be performed after study completion to test the validity of the study results. Interestingly, Lochner et al. conducted a systematic review that assessed Type II error rates in randomized trials in orthopedic trauma and concluded that, among the 117 trials included, the Type II error rate was 91% [17]. The high risk of concluding falsely negative results among orthopedic trials emphasizes the importance of a critical appraisal of study power.

C—comparison	Relative to what should the intervention be compared with? This could be another type of intervention, current standard practice, placebo, or no intervention at all
O—outcome	Define what measurements to use for measuring the effects of the intervention and the comparison. The outcome could be a direct result of the intervention/comparison and may also include side effects. Define primary and perhaps secondary outcome measurements that are valid, feasible, and reproducible

1.4 Let Evidence-Based Research Impact Clinical Work

A well-built research question is a fundamental component of EBM. However, a relevant research question can only be created through reflection of current practice and the identification of knowledge gaps. For instance, the research question may address the presentation of a specific condition, prognosis, treatment, or diagnosis. A valuable aid for this could be to systematically identify important components of the research question identified by the acronym PICO, which most commonly is used in randomized controlled trials. The acronym is defined by patient characteristics, intervention, comparison, and outcome. The features of the PICO are summarized in Fact Box 1.3.

A specific and feasible clinical question will facilitate the next step, which is to find the best available evidence.

In combination with a well-formulated question, a precise search strategy will facilitate finding relevant literature for your research questions, in an almost innumerable amount of available scientific articles. The Medical Literature Analysis and Retrieval System Online (MEDLINE) database is considered one of the most comprehensive databases and is an excellent primary choice for health-care providers since it provides both primary and secondary literature for medicine. The identified articles should thereafter be assessed based on study design and level of evidence. Especially valuable for a busy practicing clinician are the “filtered resources,” which are highly rated in the hierarchy of evidence. Examples of filtered resources are well-conducted systematic reviews, in which a synthesis of evidence has already been performed. A systematic review has been defined by the Cochrane Collaboration as “A review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyze data from the studies that are included in the review. Statistical methods (meta-analysis) may or may not be used to analyze and summarize the results of the included studies” [12]. Although the information in the systematic

Fact Box 1.3: The PICO to Formulate Research Questions

P—patient characteristics	Define the population for which the investigation is aimed. This includes demographics, the clinical characteristics (e.g., diagnoses and conditions), and the clinical setting
I—intervention	What is the exact intervention aimed to be investigated? This includes all types of treatments and diagnostic tests

review has already been scrutinized, it is the readers' responsibility to critically evaluate the included studies and to relate them to the rationale of the systematic review and the specific question. A systematic review should be carried out based on strict, predefined inclusion and exclusion criteria for all eligible articles. The search strategy should be clearly illustrated, and it should be obvious and reproducible how the articles have been assessed for inclusion and critical quality appraisal [1, 14]. It is of importance that all articles found in the literature search are evaluated objectively and that all articles meeting eligibility criteria are included, regardless of what results the studies present. Otherwise, there is an evident risk of selection bias in the systematic review [20]. Finally, if a meta-analysis is performed, the reader should evaluate data extraction, pooling of data, and the statistical methodology for quantitative synthesis of data, including data heterogeneity assessment. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [19] has been developed to facilitate the conduction of a systematic review and could also be applied when evaluating a systematic review. In general, most journals require that the authors have applied the PRISMA statement to publish a systematic review. Other valuable resources for systematic reviews and meta-analyses are, for example, the Cochrane Database of Systematic Reviews (the Cochrane Collaboration) and the Database of Abstracts of Reviews of Effects (DARE; National Institute for Health Research).

A disadvantage of filtered resources is that it takes time to conduct such a synthesis of evidence and that more recent research therefore is at risk of being missed if solely relying on such sources of information. Moreover, primary studies related to, for example, newer conditions, interventions, or technologies are often insufficient for a conduction of a systematic review. Unfiltered resources, or primary studies, provide the most up-to-date research and can include important conclusions that might have been published since the filtered resources were published. Finding the best available evidence among primary studies demands that the clinician under-

takes an evidence-based approach toward the topic on his or her own. This requires skills of the reader that may take time and education to master. Nevertheless, to develop such skills is something that is nearly mandatory in the modern world of EBM, and increased understanding of how to independently determine study validity and applicability is encouraged. Another usable instrument in the arsenal of evaluating study quality is the strength of recommendation taxonomy (SORT) [8] system which is useful when the clinician has failed to find the requested information via randomized controlled trials, meta-analyses, or systematic reviews. The SORT system could be used alone or as a complement to the level of evidence hierarchy for quality appraisal of individual studies and yields a code ranging from A to C where A represents the best possible evidence (Fact Box 1.4).

Fact Box 1.4: The SORT Rating System

Code	Definition
A	Consistent, high-quality, patient-oriented evidence
B	Inconsistent or limited-quality patient-oriented evidence
C	Consensus, disease-oriented evidence, usual practice, expert opinion, or case series for studies of diagnosis, treatment, prevention, or screening

After undertaking a thorough review of the literature and appreciating what is the best available evidence, the most important thing for a clinician is to subsequently incorporate this into clinical practice. This entails a combination of having the actual requisites to provide such health-care and also to ascertain that the patient's needs and preferences are incorporated in the algorithm of treatment. A complete utilization of EBM relays on a patient-centered health-care as well as a scientific approach to treatment. This means that the clinician has a responsibility to function as the expert of the area, presenting evidence-based alternatives to treatment, while the ultimate decision is characterized by a shared decision-making between the clinician and the patient.

1.5 The Future for EBM

Gordon Guyatt and Benjamin Djulbegovic recently published an article in *The Lancet* that reviewed the historical progress in EBM and highlighted the future directions of EBM [7]. The authors emphasized that the EBM movement has led to several related initiatives, such as measurements of the quality of care, registration of trials, improved publishing standards, discontinuation of inaccurate interventions in clinical practice, and identification of over- or underdiagnosis and over- or undertreatment. However, they also emphasized that EBM still has several future challenges to face and that EBM needs to evolve [7]. One main challenge is to ensure rapid and efficient production of systematic reviews and practice guidelines, matching the pace of which new evidence is published today. This ambition requires experienced research teams dedicated to produce comprehensive summaries of evidence rapidly, which would be facilitated by electronic interventions such as electronic platforms and automatized and text mining software [22]. The authors concluded that EBM must be developed alongside the technological evolution in terms of the use of different types of electronic devices to access medical records and data, as well as social media to reach out to the patients and the society. Furthermore, an established theory of health-care decision-making is yet to be generated, and the authors stress that collaboration with other disciplines, such as cognitive and decision sciences, is necessary to reach this goal. Clinicians should be provided with practical tools that facilitate shared decision-making in terms of making it both efficient and a positive experience for both clinicians and patients [7].

Progress of EBM is certain, and it will continue to be a cornerstone for providing the best possible health-care worldwide. Thus, EBM is considered as an essential part of medicine, which entails that it is required for clinicians and researchers to embrace the concept and to learn how to master it. Situations where EBM and clinical experience contrast might still occur, which stresses the importance of evaluating each case

individually and to continuously reevaluate current practice and perform further research when knowledge gaps are identified.

Take-Home Message

- The practice of EBM is characterized by three equally important fundamental principles: the best available research, the clinical experience, and the patient's perspective.
- The best available research should be evaluated while acknowledging the strengths and limitations of each study design, and evidence should be established based on the cumulative results from several types of study designs.
- There are several useful practical tools that could aid in critical appraisal of the literature.
- The concept of EBM is an essential part of the modern medicine and a continuous process, including reevaluation of current practice, to identify knowledge gaps for further research.

References

1. Aromataris E, Fernandez R, Godfrey CM, Holly C, Khalil H, Tungpunkom P. Summarizing systematic reviews: methodological development, conduct and reporting of an umbrella review approach. *Int J Evid Based Healthc.* 2015;13(3):132–40. <https://doi.org/10.1097/xeb.0000000000000055>.
2. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al. Grading quality of evidence and strength of recommendations. *BMJ.* 2004;328(7454):1490. <https://doi.org/10.1136/bmj.328.7454.1490>.
3. Bhandari M, Giannoudis PV. Evidence-based medicine: what it is and what it is not. *Injury.* 2006;37(4):302–6. <https://doi.org/10.1016/j.injury.2006.01.034>.
4. Bhandari M, Jain AK. The need for evidence-based orthopedics. *Indian J Orthop.* 2007;41(1):3. <https://doi.org/10.4103/0019-5413.30517>.
5. Bhandari M, Swiontkowski MF, Einhorn TA, Tornetta P 3rd, Schemitsch EH, Leece P, et al. Interobserver agreement in the application of levels of evidence to scientific papers in the American volume of the *Journal of Bone and Joint Surgery.* *J Bone Joint Surg Am.* 2004;86-a(8):1717–20.
6. Codman EA. The classic: a study in hospital efficiency: as demonstrated by the case report of first five years of private hospital. *Clin Orthop Relat Res.* 2013;471(6):1778–83. <https://doi.org/10.1007/s11999-012-2751-3>.
7. Djulbegovic B, Guyatt GH. Progress in evidence-based medicine: a quarter century on. *Lancet.*

- 2017;390(10092):415–23. [https://doi.org/10.1016/s0140-6736\(16\)31592-6](https://doi.org/10.1016/s0140-6736(16)31592-6).
8. Ebell MH, Siwek J, Weiss BD, Woolf SH, Susman J, Ewigman B, et al. Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature. *J Am Board Fam Pract.* 2004;17(1):59–67.
 9. Evans D. Hierarchy of evidence: a framework for ranking evidence evaluating healthcare interventions. *J Clin Nurs.* 2003;12(1):77–84.
 10. Evidence-Based Medicine Working Group. Evidence-based medicine. A new approach to teaching the practice of medicine. *JAMA.* 1992;268(17):2420–5.
 11. Grant HM, Tjoumakaris FP, Maltenfort MG, Freedman KB. Levels of evidence in the clinical sports medicine literature: are we getting better over time? *Am J Sports Med.* 2014;42(7):1738–42. <https://doi.org/10.1177/0363546514530863>.
 12. Green SHJ. Glossary of terms in the cochrane collaboration. The cochrane collaboration. 2015. <http://www.cochrane.org/sites/default/files/uploads/glossary.pdf>.
 13. Guyatt G, Rennie D, Meade OM, Cook JD. Users' guide to the medical literature: a manual for evidence-based clinical practice. Boston, MA: McGraw-Hill; 2014.
 14. Harris JD, Quatman CE, Manring MM, Siston RA, Flanigan DC. How to write a systematic review. *Am J Sports Med.* 2014;42(11):2761–8. <https://doi.org/10.1177/0363546513497567>.
 15. Jackson R, Ameratunga S, Broad J, Connor J, Lethaby A, Robb G, et al. The GATE frame: critical appraisal with pictures. *Evid Based Med.* 2006;11(2):35–8. <https://doi.org/10.1136/ebm.11.2.35>.
 16. Lind J. A treatise of scurvy. In three parts. Containing an enquiry into the nature, causes and cure, of that disease. Together with a critical and chronological view of what has been published on the subject. Edinburgh: Printed by Sands, Murray, and Cochran; 1753.
 17. Lochner HV, Bhandari M, Tornetta P 3rd. Type-II error rates (beta errors) of randomized trials in orthopaedic trauma. *J Bone Joint Surg Am.* 2001;83-a(11):1650–5.
 18. Matthews J. Quantification and quest for medical certainty. Princeton, NY: Princeton University Press; 1995.
 19. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol.* 2009;62(10):1006–12. <https://doi.org/10.1016/j.jclinepi.2009.06.005>.
 20. Murad MH, Montori VM, Ioannidis JP, Jaeschke R, Devereaux PJ, Prasad K, et al. How to read a systematic review and meta-analysis and apply the results to patient care: users' guides to the medical literature. *JAMA.* 2014;312(2):171–9. <https://doi.org/10.1001/jama.2014.5559>.
 21. Oxman AD, Sackett DL, Guyatt GH. Users' guides to the medical literature. I. How to get started. The Evidence-Based Medicine Working Group. *JAMA.* 1993;270(17):2093–5.
 22. Paynter R, Banez LL, Berliner E, Erinoff E, Lege-Matsuura J, Potter S, et al. EPC methods: an exploration of the use of text-mining software in systematic reviews. Rockville, MD: Agency for Healthcare Research and Quality (US); 2016.
 23. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest.* 1986;89(Suppl 2):2s–3s.
 24. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ.* 1996;312(7023):71–2.
 25. Samuelsson K, Desai N, McNair E, van Eck CF, Petzold M, Fu FH, et al. Level of evidence in anterior cruciate ligament reconstruction research: a systematic review. *Am J Sports Med.* 2013;41(4):924–34. <https://doi.org/10.1177/0363546512460647>.
 26. Straus ES, Glasziou P, Richardson W, Haynes R. Evidence-based medicine. How to practice and teach EBM. Edinburgh: Churchill Livingstone; 2005.
 27. Watts G. Let's pension off the "major breakthrough". *BMJ.* 2007;334(Suppl 1):s4. <https://doi.org/10.1136/bmj.39034.682778.94>.
 28. Wright JG, Swiontkowski MF, Heckman JD. Introducing levels of evidence to the journal. *J Bone Joint Surg Am.* 2003;85-a(1):1–3.



What Is the Hierarchy of Clinical Evidence?

2

Vishal S. Desai, Christopher L. Camp,
and Aaron J. Krych

2.1 Introduction: Why Do We Need a Hierarchy?

Not all evidence is created equal. In the age of information, when data and results can be disseminated within seconds, it remains ever-important to ensure that evidence is accurately organized and graded. In order for clinical research to be translated into improving the quality of patient care and provide practice-changing evidence, a researcher/provider must be able to efficiently distinguish the quality and application of the information they are reading [22]. From the researcher’s perspective, an organized hierarchy allows him or her to select the optimal study design for their clinical question, as will be demonstrated later in this chapter. And from the reader’s perspective, it allows for results to be interpreted in a clinically appropriate context and be compared on the basis of strength relative to the results of other investigators [41]. Selection of the appropriate study type can serve to minimize study bias and improve quality, validity, and reliability. This minimization in bias ultimately has the potential to translate into improved clinical decision-making [21]. There is a growing consensus for the importance of effective quality

and strength stratification of published literature. Many journals have now begun requiring authors to declare a level of evidence for their submitted work [48].

2.2 The Hierarchy of Evidence and Study Selection

The hierarchy of clinical evidence is traditionally introduced as a pyramid that allows the reader to organize study design by the level of evidence it yields. Figure 2.1 is a commonly seen rendering of this important principle. This basic framework has been presented and corroborated by several authors since initially appearing in the literature as early as 1979 [19, 23, 40]. Primary, or unfiltered, evidence involves the direct observation and interpretation of data. Secondary, or filtered, evidence relies on the reorganization and interpretation of data that has already been published in a primary study in an attempt to efficiently draw conclusions based on larger sums of data. While the hierarchy does serve to delineate “primary” from “secondary” evidence and visually illustrates the relative strength of study types [1], it does not address how to apply a particular study design to the clinical question of the investigator. Use of the “PICO framework” (which stands for **P**atient, **P**roblem, **P**opulation; **I**ntervention; **C**omparison, **C**ontrol; **O**utcome) may guide development of an appropriate clinical

V. S. Desai · C. L. Camp · A. J. Krych (✉)
Department of Orthopedic Surgery and Sports
Medicine, Mayo Clinic, Rochester, MN, USA
e-mail: Desai.vishal@mayo.edu; Camp.christopher@mayo.edu; krych.aaron@mayo.edu

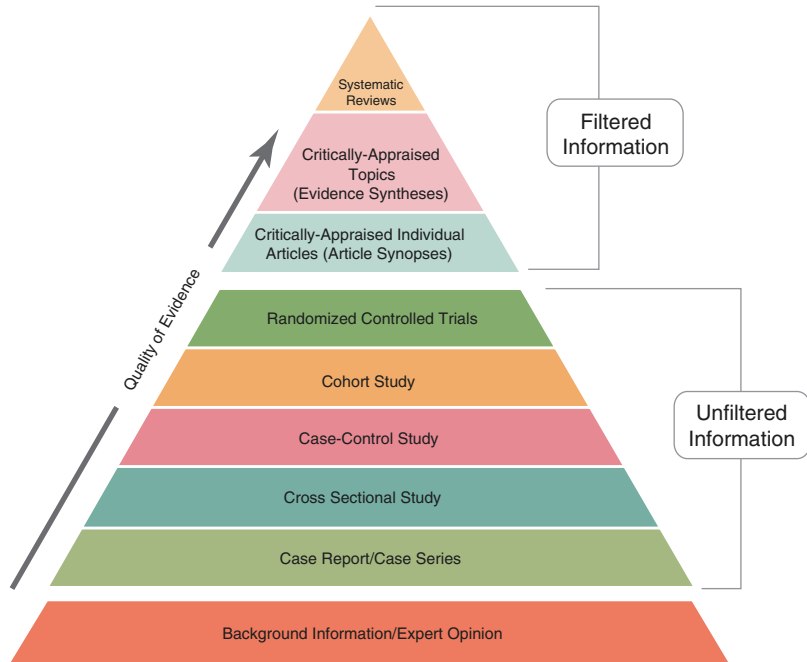
question to which a study design can later be tailored. This model serves to clearly outline key variables of the question so that subsequent data collection and analysis can be performed in a focused manner [26, 43] (Fig. 2.2: **P**atient, **P**roblem, **P**opulation; **I**ntervention; **C**omparison, **C**ontrol; **O**utcome). Focused around the topic of anterior cruciate ligament (ACL) reconstruction outcomes, we will introduce and parse out the hierarchy of clinical evidence as well as demonstrate how its components can be applied differently to answer clinical questions. Let us begin first with a discussion of primary (unfiltered) research including expert opinion, case reports, cross-sectional study, case-control study, cohort study, and randomized controlled trials.

2.3 Primary/Unfiltered Research

2.3.1 Expert Opinion

Prior to popularization of evidence-based medicine and peer review, expert opinion was the predominant form of “clinical evidence” used to communicate findings to a larger audience. In fact, a significant portion of orthopedic literature that has been influencing therapeutic decision-making of surgeons for many years is based in expert opinion and case series [18]. As it is reliant on the observations and experiences of an individual (or group) and may lack a reproducible or methodical design, some argue that it does not belong in the hierarchy of evidence at all [14]. To reduce the dependence of

Fig. 2.1 Pyramid of evidence-based medicine



P	I	C	O
Patient Population Problem	Intervention	Comparison	Outcome

Fig. 2.2 The PICO format allows the examiner to methodically develop all components of a complete clinical question

the results on simply the observations of an individual, expert opinion can be strengthened by pooling together and appropriately summarizing the opinions of multiple, well-respected members of a scientific community. Consider the following by Zantop et al. [49] in which 20 panelists from around the world were surveyed on their preferential technique and rehabilitation protocol for ACL reconstruction. The purpose of the study was to summarize the collective practices of this panel of experts to inform clinicians of the current technical conventions predominantly used in ACL reconstruction. The authors defined “expert” as those who were identified in the specific subject matter to have contributed to the National Library of Medicine or present clinical evidence at international meetings [49]. Expert opinions provide an efficient way for the clinical community at large to be updated on the current trends and pitfalls of practice in the community at large. However, based on the design of expert opinion articles, it becomes clear that panelist bias, small sample sizes, and lack of a rigorous data collection protocols limit their overall impact and relative significance [14, 47, 49]. Despite their shortcomings, the information presented in a high-quality, thorough expert opinions may serve to have more practical utility than that of a poorly designed clinical study, and therefore, they will remain a crucial component of the medical literature [47].

2.3.2 Case Report

Case reports document events or outcomes of individual or small groups of patients that the author finds to be of significance. Whether or not the presented findings are positive or negative, they may be useful in dictating future patient care [14]. Huth outlined four major applications of case reports in his discussion on medical literature [27]:

1. A unique case that may represent a previously unknown syndrome or disease
2. A case with the previously unreported association of two distinct diseases, suggesting a possible relationship between them

3. An “outlier” with features strikingly outside the realm of what is usually seen with a particular disease
4. An unexpected response or course suggesting a previously unrecognized therapeutic or adverse effect of intervention

In the realm of orthopedic surgery, surgical complications may be of significance. Although surgical complications remain a traditionally underreported topic in orthopedic literature, case reports may serve to inform surgeons of potential risks and how to avoid them. For example, a case report by Heng et al. illustrated a 35-year-old man who developed a peri-ACL, distal femoral fracture running through the two femoral tunnels following double-bundle ACL reconstruction. Furthermore, the authors go on to identify this case as the first of its kind following a thorough literature review [24]. On the one hand, a clinician, after having read this, may refine surgical technique or be more cognizant of specific risk factors in his/her patients that may predispose them to a similar outcome. However, although the authors did allude to a possible mechanism of this complication, without an appropriately conducted scientific experiment, the basis of their commentary largely remains conjecture. Case reports may inspire a testable hypothesis for future studies to execute and be successful in providing short and succinct learning points for the reader [17]; however, they provide limited value when attempting to investigate and answer specific clinical questions. Further limitations of case reports include a lack of generalizability, the inability to establish cause-effect relationships, publication bias, and the risk of “overinterpreting” the findings beyond their simple anecdotal intent [38].

2.3.3 Cross-sectional Study

The cross-sectional study attempts to answer the question, “what is happening right now?” One of the most common applications of the cross-sectional study is in determining the prevalence of a condition or diagnosis at a particular time.

Let us return to our theme of ACL reconstruction and consider the following cross-sectional study. Farber et al. [16] investigated, “Which management strategies for ACL injury are currently the predominant forms of therapy applied by team physicians in Major League Soccer (MLS)?” The authors selected a focused question that attempts to answer what is currently happening rather than trying to establish a temporal, causative, or correlative relationship of any sort. In order to answer this question, the investigators sent out a survey which required respondents to provide their particular surgical technique when treating MLS players [16]. It is important to note that while this study may have employed a survey-based model just as the expert opinion example did, they can be differentiated by how participants were selected and the types of questions they are trying to answer. Rather than a more loosely defined group of “experts,” cross-sectional studies identify a specific group, in this case MLS team physicians, and select a specific question for them to answer. Their responses can then be statistically pooled and homogeneously formatted. The primary shortcoming of cross-sectional studies is their inability to generate temporal or causative relationships between exposures and outcomes. In our example case, although the study was able to identify which reconstruction techniques are currently prevalent among MLS team physicians, it does not begin to address the relative outcomes of the patients who underwent these interventions nor the risk factors that may have placed them at risk for the injury in the first place. Therefore, while they do not necessarily serve to show trends or relationships between multiple variables, cross sections are effective for gathering and presenting large volumes of generalizable data.

2.3.4 Case-Control Study

Case-control studies begin with a sample of patients in whom a selected outcome has already occurred and been identified. Patients with and without this outcome are then compared for dif-

ferences in exposures and risk factors between the two groups. Because case-control studies begin with the outcome, they are always inherently retrospective in design. In addition, because the patient outcome is selected by the investigator, case controls serve as the ideal study design for analyzing root causes and risk factors of rare diseases. For example, you may have identified a group of your patients who failed single-bundle (SB) ACL reconstruction. And as you think about why this may have happened, the question arises, “what predisposed these patients to failing surgery?” A well-designed case control, such as the following, may be suited to answer this question. Parkinson et al. [39] identified 123 SB ACL reconstruction patients with long-term follow-up who met their inclusion criteria. Patients who demonstrated failure at 2-year follow-up had higher rates of medial and lateral meniscus deficiency, shallower femoral tunnel positioning, and younger patient age. The authors were able to conclude that medial meniscus deficiency was the most significant factor to predict graft failure in SB anatomic ACL reconstructions [39] (Fact Box 2.1: Odds Ratio). While they serve a very important role in clinical research, case-control studies are not without limitation. In particular, they are prone to recall bias in which there are discrepancies in the accuracy and thoroughness by which prior exposures are recalled among study participants [30]. For example, in a study of risk factors for chronic ACL deficiency, participants with ACL pathology are likely to more thoroughly search their memories for injuries and underlying causes compared to their unaffected counterparts, thus creating bias in the collected data. Another important shortcoming of case-control studies is their limited ability to demonstrate causation. As causation is typically best accomplished in prospective studies, the retrospective nature of a case control makes it a poor indicator for causal inference between exposure and outcome. In the case of our above example, the study we highlighted more effectively showed a correlation between medial meniscus deficiency and ACL reconstruction failure rather than showing that medial meniscus deficiency caused a failure in ACL reconstruction.

Fact Box 2.1: Odds Ratio

Odds ratio (OR) is a commonly applied statistical tool in case-control studies that illustrates how strongly a given characteristic of a subject is associated with the presence or absence of another characteristic of that subject. In the case of a study examining risk factors for a disease, it answers the question, “what are the odds that those who developed the disease were exposed to the risk factor compared to the odds that those who developed the disease were *NOT* exposed to the risk factor?”

2.3.5 Cohort Study

Whereas a case control seeks to answer the question of “what happened that caused this outcome?”, cohort studies serve to answer the question of “what will happen?” Instead of controlling the intervention, the researcher instead observes the outcomes in the sample being studied. This makes cohort studies suitable for studying the natural progression of exposures, diseases, and risk factors [4]. Unlike case-control studies, cohorts are unique in that they can be executed both prospectively and retrospectively [4]. We have compared and contrasted two cohort studies to illustrate this principle.

First, let us consider the following prospective cohort in which the examiner collects patient exposure information at the beginning of the study and then analyzes the outcomes at a time after that [4]. *In 2014, the Multicenter ACL Revision Study (MARS) Group enrolled 1205 patients to understand the effect that graft choice has on outcomes in patients undergoing revision ACL reconstruction (ACLR). They hypothesized that an autograft reconstruction would result in increased sports function and activity levels and decreased osteoarthritis symptoms and graft failure rates compared to allograft reconstruction. Baseline questionnaires and function scores were measured in all patients; after following these patients for a minimum of 2 years, these same*

questionnaires and function scores were readministered, and outcomes between those who underwent an autograft and allograft revision ACLR were compared [13]. In other words, the authors selected a group of patients with certain exposures (autograft vs allograft), followed them over a course of time, and finally evaluated them for how their long-term outcomes compared.

In retrospective cohorts, however, the examiner begins by collecting patient exposure information from some time in the past and then analyzes the outcomes that developed between that time and the present time [4]. Kim et al. compared functional outcomes and stability of ACL reconstructions of double-bundle versus single-bundle ACL remnant pullout suture techniques using this study design. The examiners selected 44 patients who underwent single-bundle reconstruction with remnant tensioning (Group 1) and 56 patients who underwent double-bundle reconstruction (Group 2) 5–8 years prior to the beginning of the study; then, outcomes were measured at a minimum of 3 years after the surgical date, and no statistically significant difference was identified between the two groups [29]. Their question sought to determine how two groups of patients, treated by different interventional techniques in the past (exposure), progressed over a period of time and how their outcomes compared; thus, a retrospective cohort proved to be the optimal choice.

Either of the above two orientations may serve to provide concrete outcome data between two groups of patients (Fact Box 2.2: **Relative Risk**). In the context of orthopedic surgery, cohorts are especially useful in providing strong, clinically validated support for the effectiveness of surgical interventions. Although cohort studies may be cumbersome and costly to execute, especially when prospective, they have the benefit of providing a temporal sequence of events between exposure and outcome and, as a result, render compelling, clinically useful findings including potential causation. In fact, it has been suggested that a rigorously designed cohort study has the potential to yield results similar to that of the more powerful randomized controlled trial [4].

Fact Box 2.2: Relative Risk

Relative risk (RR), similar to the odds ratio, is a statistical tool more commonly applied in cohort studies and represents the compared probabilities of an outcome occurring in the exposed versus unexposed group. In the case of a study examining outcomes of certain risk factors, it answers the question, “what is the risk of developing a disease in the group who was exposed to a risk compared to the risk of developing the same outcome in a group who was NOT exposed to the risk factor?” Although the distinction between RR and OR may appear subtle, it is nonetheless a crucial one.

2.3.6 Randomized Controlled Trial (RCT)

RCTs are among the most robust experimental study designs and have been described in the past as the “gold standard” of evidence-based medicine [45]. Examiners place the study subjects in either a control or intervention group; thus, because the examiner is in the position to regulate the exposure or interventional strategy in question, RCTs are inherently prospective [14]. A quintessential RCT may be organized to answer the question of whether or not a difference exists between an exposure and control (or an alternative exposure) group when compared prospectively. *Let us consider the following RCT by Mayr et al. [32]. The authors sought to answer the question: “if a group of patients in need of undergoing ACL reconstruction were randomized to receive either a single-bundle technique or double-bundle technique, how would their outcomes differ at long-term follow-up?” Sixty-four patients were randomly and evenly split to undergo one of the two surgical techniques. At a minimum of 2 years after the operation, patients were both subjectively and objectively evaluated for knee function. Their results showed no statistically significant difference in outcomes between the two interventional groups, implying a relative*

clinical equivalency between the two techniques [32]. Due to the deliberate and focused design of RCTs, it has been suggested that they can demonstrate causality between the exposure and outcome more so than other study designs can [33].

As the name implies, randomization is one of the most important features of RCTs. By “randomly” placing subjects in different experimental groups, the examiner is able to not only decrease potentially biasing factors but also improve the internal validity of the study [45]. When random allocation is performed properly, the participants that are assigned to the treatment and control arms of a study are placed there by chance and no preference of the investigator; this minimizes the burden that known and unknown patient factors may have on the results of the study and provides the highest quality of data. Although there are many different protocols that can be applied to carry out effective randomization, as discussed by Kim et al., they all attempt to achieve this same basic objective of study design [28].

When discussing RCTs, an understanding of the “intention-to-treat (ITT)” principle is essential. Simply put, it states that regardless of patient noncompliance, deviations from the experimental protocol, or patient withdrawal from a study, a patient should remain a member of the original group, whether it be control or experimental, that he/she was assigned to [20, 36]. Why is this significant? ITT can serve to minimize bias and improve the validity of the data presented by the examiner in the event of patient noncompliance or patient withdrawal [20]. Let us illustrate this principle with an example using our recurring theme of ACL reconstruction to further elucidate the concept. *Consider, for example, a group of 100 patients who have suffered ACL injury. Of these, 50 are randomly assigned to undergo surgical reconstruction (experimental) whereas the other 50 are assigned conservative treatment and physical therapy (control). Patients are consented, and the plan is to evaluate the patients 2 years from now to see how both groups fared over that period of time. However, during the experimental period, ten patients from the control group feel that they are not improving and*

decide to seek surgical intervention through another avenue. What now are we to do with this data which has been altered by patient movement between our groups? The ITT principle contends that patients should be analyzed in the group to which they were initially randomized. If the ten patients who left the control group are excluded from its analysis, the remaining 40 patients in that group now represent a clinically healthier and skewed sample of patients thus biasing the results. In other words, patients who are faring more poorly in the control group are more likely to abandon it. The Spine Patient Outcomes Research Trial (SPORT) of 2006 beautifully illustrates a real-life example of this dilemma. A group of 501 patients who were all surgical candidates for lumbar discectomy for intervertebral disk herniation were randomized to either undergo standard open discectomy or non-operative treatment for their persistent signs and symptoms. Although patients were initially randomized evenly between the two groups, only 50% of patients assigned to the surgery group had undergone surgery at 3 months; in this same 3-month period, 30% of those assigned to the non-operative arm of the study underwent surgery. The examiners used ITT analysis on the data which did show small and not statistically significant improvement in the surgical group over the nonsurgical. However, the substantial crossover of patients between the two groups led the author to conclude that superior or equivalence of treatments could not be established and thus illustrated a key limitation of executing RCTs [46]. If all patients are analyzed in the groups to which they are initially assigned, as per ITT analysis, examiners can improve protocol adherence and decrease bias. As a result of implementing this principle, the examiner's findings tend to become more conservative as non-compliant and withdrawn patients potentially dilute positive data [20]. ITT is not without limitations, and critics have argued that it may fail to provide an accurate picture of the data if large portions of patients change groups [25, 44]. An alternative to ITT is "as-treated" analysis in which subjects are compared with the treatment regimen that they have received regardless of

which group they were originally assigned to [15]. The SPORT trial illustrates side by side how these two analytical methods can yield drastically different results from the same patient set. In addition to the ITT analysis performed (as described above), the examiners also analyzed the data using the as-treated method which showed strong, statistically significant advantages for surgery at all follow-up times through 2 years [46]. While this does go to show that ITT can underestimate the true effect of surgery in certain cases, one clear limitation of as-treated analysis is that it does not provide the same protection against confounding variables that can be achieved by following through with the original randomization protocol [46].

Despite representing the highest level of primary evidence, RCTs are not without shortfalls. A discussion of the limitations of RCTs would be incomplete without an understanding of "clinical equipoise." This principle states that the examiner and medical community at large genuinely cannot definitely say that one arm of an experiment is superior to the other. Ideally, RCTs performed under the assumption of clinical equipoise uphold a higher ethical benchmark as they prevent the deliberate withholding of a treatment from patients which is otherwise known to be superior to the control arm of the study [7, 37]. Equipoise is necessary to ethically justify the use of randomization to select which treatment (or lack thereof) a patient will receive in a situation where the optimal treatment method is not yet agreed upon by the expert community [12]. This concept, in the context of orthopedic surgery, can be well illustrated with a discussion of sham surgery. Sham surgeries are analogous to "placebo drugs" in that they create the illusion to a patient that he or she is being treated; however, they in fact omit the therapeutically significant step in surgery [37]. On the one hand, they allow us to accurately compare the efficacy of the treatment arm; but on the other hand, they raise the possibility of denying appropriate therapy to patients of the control arm. Current literature holds that within the confines of ethical practices, equipoise, and clear, informed consent, sham surgeries do have a place in orthopedic clinical trials [12, 34].

A major challenge with RCTs lies in the expense, man power, and resources needed to perform them. In the absence of ethically managed protocols and high reporting standards, RCTs may provide erroneous information which can subsequently compromise patient care. Despite being considered the “gold standard” of study design, RCTs are not the only option to generate reliable and evidence-based results. It is recommended that if clinical questions can be answered accurately and more efficiently with a so-called “lower-impact” study design (as the ones discussed above), those avenues should be considered [6].

2.4 Secondary/Filtered Research

Secondary evidence primarily takes the form of two study designs: the systematic review and the meta-analysis. Rather than searching for and interpreting data directly, secondary evidence relies upon pooling of primary evidence that is already available and organizing it in a meaningful and novel fashion. Because it represents a summary of the best available evidence on a particular topic [21], secondary evidence is traditionally classified as the highest level on the hierarchy of evidence. Despite the perceived value of secondary evidence, critics have called into question the validity of its findings and suggest that the medical community may be overvaluing their contribution compared to well-designed, primary studies [2, 3, 5, 8]. The inclusion of methodologically flawed and non-randomized trials into systematic reviews, due to the unavailability of high-quality RCTs in the orthopedic literature, has been a recent topic of discussion [5]. The findings from a systematic review or meta-analysis are only as strong as the studies that comprise it; thus, the inclusion of low-quality studies compromises the validity of reported results. The inclusion of clinically irrelevant studies and “old” information has weakened the quality of conclusions, if any, that can be drawn from many meta-analyses. As a result, they do little to advance the standard of care or best practices [8]. Audigé et al. reviewed the

orthopedic literature and a preponderance of systematic reviews substantiated by nonrandomized and heterogeneous studies; they recommended a degree of caution when interpreting findings of secondary evidence [2].

2.4.1 Systematic Review

Systematic reviews offer a comprehensive collection of all literature available on a particular topic followed by a discussion and analysis of findings. They allow the examiner to answer the question, “What does the current literature, as a whole, say about my question?” Returning to our theme of ACL reconstruction, consider the following. *In assessing how outcomes of ACL reconstruction surgery differ between male and female patients, Ryan et al. performed a systematic review of literature and found 13 studies that met their inclusion criteria. Stratifying their findings based on “graft failure risk,” “contralateral ACL injury risk,” “knee laxity,” and “patient-reported outcomes,” their review found no evidence of clinically important difference in outcomes between male and female patients [42].* While systematic reviews do allow for a simplified, qualitative discussion of large volumes of data, what they generally lack is a head-to-head statistical analysis. Another limitation of systematic reviews is that while they do require an exhaustive search of literature, the inclusion criteria of studies that ultimately make it to the analysis stage can be doctored to the point that you now portray a biased perspective on an otherwise generalizable question. In other words, the author has the freedom to alter the inclusion criteria to include (or exclude) studies that may reinforce their argument.

2.4.2 Systematic Review with Meta-analysis

Like a systematic review, meta-analysis also begins with a comprehensive collection of literature pertinent to the examiner’s question. However, in order to be classified as meta-

analysis, the examiner must be able to pool statistics [14]. Therefore, meta-analysis requires a relative degree of homogeneity in data. When studies are combined in a meta-analysis, they are generally done using either a “fixed effects” model or “random effects” model. In the Mantel-Haenszel fixed model, the examiner assumes that one true effect size underlies all studies in the analysis and that any differences among the included studies are due to chance [31]. In the DerSimonian and Laird random effects model, the degree of heterogeneity between studies is considered to render a fixed effect model implausible and thus is the optimal selection in evaluating heterogeneity between studies [10].

Because data on the same subject matter may be presented using different metrics, units, outcome scores, etc., meta-analysis is more difficult to execute. Say one were interested in finding a comprehensive analysis comparing outcomes scores and physical exam findings for all patients who underwent single-bundle versus double-bundle ACL reconstruction that are documented in the literature. *A meta-analysis by Desai et al. identified 970 patients from 15 studies which met inclusion criteria and allowed the examiner to compare a homogenized data set. Their results showed a statistical improvement in knee kinematics of the double-bundle technique but no significant improvement over single-bundle reconstruction on clinically meaningful outcomes [11].* Just as in systematic reviews, meta-analysis is also limited by its vulnerability to bias based on the author’s decision to include or exclude studies.

2.4.3 PRISMA

How does one go about conducting an exhaustive review of literature? In 2009, Moher et al. released the “preferred reporting items for systematic reviews and meta-analyses” (PRISMA) statement. The article outlines a meticulous methodology by which to design and execute systematic reviews and meta-analyses with the intent of improving and standardizing the quality of secondary evidence available in the literature [35]. Researchers preparing a manuscript of this

design are highly encouraged to review and adhere to the PRISMA guidelines prior to beginning. To see an outline of how to systematically filter the literature to isolate the studies to more thoroughly examine for inclusion in a study, please refer to the Additional Resources section.

2.5 Levels of Evidence

Although the “levels of evidence” will be discussed at length in Chap. 5, we felt they merited a brief mention here as they are modeled off of a similar structure as the hierarchy of evidence. Figure 2.3 adapted from an editorial by Wright et al. in the *Journal of Bone and Joint Surgery* elegantly outlines the levels of evidence and the study designs that underlie them [48]. Not only do they demonstrate the relative strength of data generated by different types of studies; they serve to uniformly grade the quality recommendations derived from that data.

2.6 Trends in the Evolution of Orthopedic Literature

Over the past several decades, orthopedic literature has continued to change in both quantity and quality. Cunningham et al. [9] published a 2013 study in which they reviewed literature in eight of the highest impact journals in orthopedic medicine over the past 10 years to identify trends in the volume and level of evidence that has been published. In short, they found that the annual volume of published studies continuously increased between the years 2000 and 2010 and the number of Level I and Level II studies significantly increased during that same period. However, despite these increases in volume, they reported that much of the literature still contains Level III and Level IV studies and advocated for researchers to publish the highest level of evidence available for a given question [9] (Fig. 2.4). These findings were later corroborated in a study by Grant et al. [18] who also identified an increase in the volume of Level I and Level II evidence studies. In addition, they highlighted that the

Levels of Evidence for Primary Research Question

	Therapeutic Studies – Investigating the Results of Treatment	Prognostic Studies – Investigating the Outcome of Disease	Diagnostic Studies – Investigating a Diagnostic Test
Level I	1. Randomized controlled trial a. Significant difference b. No significant difference but narrow confidence intervals 2. Systematic review ² of Level-I randomized controlled trials (studies were homogeneous)	1. Prospective study ¹ 2. Systematic review ² of Level-I studies	1. Testing of previously developed diagnostic criteria in series of consecutive patients (with universally applied reference “gold” standard) 2. Systematic review ² of Level-I studies
Level II	1. Prospective cohort study ³ 2. Poor-quality randomized controlled trial (e.g., 80% follow-up) 3. Systematic review ² a. Level-II studies b. nonhomogeneous Level-I studies	1. Retrospective study ⁴ 2. Study of untreated controls from a previous randomized controlled trial 3. Systematic review ² of Level-II studies	1. Development of diagnostic criteria on basis of consecutive patients (with universally applied reference “gold” standard) 2. Systematic review ² of Level-II studies
Level III	1. Case-control study ⁵ 2. Retrospective cohort study ⁴ 3. Systematic review ² of Level-III studies	Case series	1. Study of nonconsecutive patients (no consistently applied reference “gold” standard) 2. Systematic review ² of Level-III studies
Level IV	Case series (no, or historical, control group)		1. Case-control study 2. Poor reference standard
Level V	Expert opinion	Expert opinion	Expert opinion

Figure adapted from The Journal of Bone & Joint Surgery: 85-A(1), 2003

Fig. 2.3 This table outlines the clinical levels of evidence and the study designs that underlie them in the context of therapeutic, prognostic, and diagnostic studies [48]

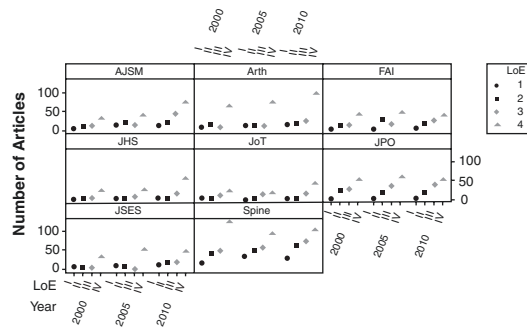
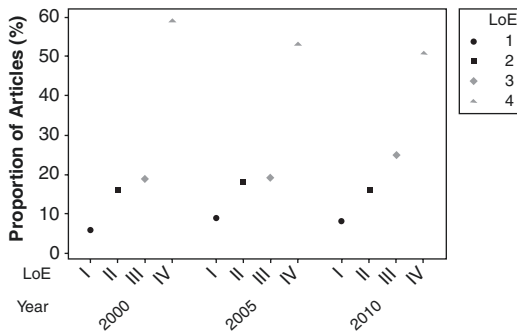


Fig. 2.4 Despite an increase in the volume of orthopedic literature published on an annual basis over the past 20 years, a high proportion of the literature is still comprised of Level III and Level IV studies [9]

largest increase in volume has been seen in diagnostic studies, compared to more modest increases in prognostic and therapeutic ones [18].

Take-Home Message

- Despite this hierarchical construct that has been developed to organize evidence based on strength, the most important factor to consider when designing a study is “what is the question that I want to answer?”
- A seemingly “stronger” study may add no additional benefit to one’s objective if it does a

poor job in answering the specific question that the examiner sought to address in the first place.

- While RCTs may represent the gold standard of clinical evidence, they may not always be the ideal choice for a particular question.
- While we encourage publishing the highest quality of evidence feasible, we stress the importance of ensuring that every scientific study begins with a well-crafted clinical question and only then should a study design be implemented.

2.7 Additional Resources and Websites

- PRISMA Flow Diagram: <http://prismastatement.org/documents/PRISMA%202009%20flow%20diagram.pdf>.
- University of Central Florida College of Medicine Harriet F. Ginsburg Health Sciences Library Evidence-Based Medicine Guide: <http://guides.med.ucf.edu/c.php?g=174363&p=1149202>.

References

1. Audige L, Ayeni OR, Bhandari M, Boyle BW, Briggs KK, Chan K, Chaney-Barclay K. A practical guide to research: design, execution, and publication. *Arthroscopy*. 2011;27(4 Suppl):S1–112. <https://doi.org/10.1016/j.arthro.2011.02.001>.
2. Audige L, Bhandari M, Griffin D, Middleton P, Reeves BC. Systematic reviews of nonrandomized clinical studies in the orthopaedic literature. *Clin Orthop Relat Res*. 2004;427:249–57.
3. Bhandari M, Morrow F, Kulkarni AV, Tornetta P 3rd. Meta-analyses in orthopaedic surgery. A systematic review of their methodologies. *J Bone Joint Surg Am*. 2001;83-a(1):15–24.
4. Bryant DM, Willits K, Hanson BP. Principles of designing a cohort study in orthopaedics. *J Bone Joint Surg Am*. 2009;91(Suppl 3):10–4. <https://doi.org/10.2106/jbjs.h.01597>.
5. Chaudhry H, Mundi R, Singh I, Einhorn TA, Bhandari M. How good is the orthopaedic literature? *Indian J Orthop*. 2008;42(2):144–9. <https://doi.org/10.4103/0019-5413.40250>.
6. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000;342(25):1887–92.
7. Cook C, Sheets C. Clinical equipoise and personal equipoise: two necessary ingredients for reducing bias in manual therapy trials. *J Man Manip Ther*. 2011;19(1):55–7. <https://doi.org/10.1179/106698111x12899036752014>.
8. Court-Brown CM, McQueen MM. How useful are meta-analyses in orthopedic trauma? *J Trauma*. 2011;71(5):1395–9. <https://doi.org/10.1097/TA.0b013e318208f983>.
9. Cunningham BP, Harmsen S, Kweon C, Patterson J, Waldrop R, McLaren A, McLemore R. Have levels of evidence improved the quality of orthopaedic research? *Clin Orthop Relat Res*. 2013;471(11):3679–86. <https://doi.org/10.1007/s11999-013-3159-4>.
10. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7(3):177–88.
11. Desai N, Bjornsson H, Musahl V, Bhandari M, Petzold M, Fu FH, Samuelsson K. Anatomic single- versus double-bundle ACL reconstruction: a meta-analysis. *Knee Surg Sports Traumatol Arthrosc*. 2014;22(5):1009–23. <https://doi.org/10.1007/s00167-013-2811-6>.
12. Dowrick AS, Bhandari M. Ethical issues in the design of randomized trials: to sham or not to sham. *J Bone Joint Surg Am*. 2012;94(Suppl 1):7–10. <https://doi.org/10.2106/jbjs.1.00298>.
13. MARS Group. Effect of graft choice on the outcome of revision anterior cruciate ligament reconstruction in the Multicenter ACL Revision Study (MARS) Cohort. *Am J Sports Med*. 2014;42(10):2301–10. <https://doi.org/10.1177/0363546514549005>.
14. Elamin MB, Montori VM. The hierarchy of evidence: from unsystematic clinical observations to systematic reviews. In: Burneo JG, Demaerschalk BM, Jenkins ME, editors. *Neurology: an evidence-based approach*. New York, NY: Springer; 2012. p. 11–24.
15. Ellenberg JH. Intent-to-treat analysis versus as-treated analysis. *Drug Inf J*. 1996;30(2):535–44. <https://doi.org/10.1177/009286159603000229>.
16. Farber J, Harris JD, Kolstad K, McCulloch PC. Treatment of anterior cruciate ligament injuries by major league soccer team physicians. *Orthop J Sports Med*. 2014;2(11):2325967114559892. <https://doi.org/10.1177/2325967114559892>.
17. Gopikrishna V. A report on case reports. *J Conserv Dent*. 2010;13(4):265–71. <https://doi.org/10.4103/0972-0707.73375>.
18. Grant HM, Tjoumakaris FP, Maltenfort MG, Freedman KB. Levels of evidence in the clinical sports medicine literature: are we getting better over time? *Am J Sports Med*. 2014;42(7):1738–42. <https://doi.org/10.1177/0363546514530863>.
19. Greenhalgh T. How to read a paper. Getting your bearings (deciding what the paper is about). *BMJ*. 1997;315(7102):243–6.
20. Gupta SK. Intention-to-treat concept: a review. *Perspect Clin Res*. 2011;2(3):109–12. <https://doi.org/10.4103/2229-3485.83221>.
21. Guyatt G, Rennie D, Meade M, Cook D. *Users' guides to the medical literature: a manual for evidence-based clinical practice*. Chicago, IL: AMA Press; 2002.
22. Guyatt GH, Haynes RB, Jaeschke RZ, Cook DJ, Green L, Naylor CD, Wilson MC. *Users' Guides to the Medical Literature: XXV. Evidence-based medicine: principles for applying the Users' Guides to patient care. Evidence-Based Medicine Working Group*. *JAMA*. 2000;284(10):1290–6.
23. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. *Users' guides to the medical literature. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group*. *JAMA*. 1995;274(22):1800–4.
24. Heng CH, Wang Bde H, Chang PC. Distal femoral fracture after double-bundle anterior cruciate ligament reconstruction surgery. *Am J Sports Med*. 2015;43(4):953–6. <https://doi.org/10.1177/0363546514563908>.

25. Hollis S, Campbell F. What is meant by intention to treat analysis? Survey of published randomised controlled trials. *BMJ (Clin Res ed)*. 1999;319(7211):670–4.
26. Huang X, Lin J, Demner-Fushman D. Evaluation of PICO as a knowledge representation for clinical questions. *AMIA Ann Symp Proc*. 2006;2006:359–63.
27. Huth EJ. Writing and publishing in medicine. Baltimore: Lippincott Williams and Wilkins; 1999. p. 103–10.
28. Kim J, Shin W. How to do random allocation (randomization). *Clin Orthop Surg*. 2014;6(1):103–9. <https://doi.org/10.4055/cios.2014.6.1.103>.
29. Kim SH, Jung YB, Song MK, Lee SH, Jung HJ, Lee HJ, Jung HS. Comparison of double-bundle anterior cruciate ligament (ACL) reconstruction and single-bundle reconstruction with remnant pull-out suture. *Knee Surg Sports Traumatol Arthrosc*. 2014;22(9):2085–93. <https://doi.org/10.1007/s00167-013-2619-4>.
30. Kopec JA, Esdaile JM. Bias in case-control studies. A review. *J Epidemiol Community Health*. 1990;44(3):179–86.
31. Mantel N, Brown C, Byar DP. Tests for homogeneity of effect in an epidemiologic investigation. *Am J Epidemiol*. 1977;106(2):125–9.
32. Mayr HO, Benecke P, Hoell A, Schmitt-Sody M, Bernstein A, Suedkamp NP, Stoehr A. Single-bundle versus double-bundle anterior cruciate ligament reconstruction: a comparative 2-year follow-up. *Arthroscopy*. 2016;32(1):34–42. <https://doi.org/10.1016/j.arthro.2015.06.029>.
33. McCarthy CM. Randomized controlled trials. *Plast Reconstr Surg*. 2011;127(4):1707–12. <https://doi.org/10.1097/PRS.0b013e31820da3eb>.
34. Mehta S, Myers TG, Lonner JH, Huffman GR, Sennett BJ. The ethics of sham surgery in clinical orthopaedic research. *J Bone Joint Surg Am*. 2007;89(7):1650–3. <https://doi.org/10.2106/jbjs.f.00563>.
35. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med*. 2009;6(7):e1000097. <https://doi.org/10.1371/journal.pmed.1000097>.
36. Montori VM, Guyatt GH. Intention-to-treat principle. *CMAJ*. 2001;165(10):1339–41.
37. Mundi R. Design and execution of clinical trials in orthopaedic surgery. *Bone Joint Res*. 2014;3(5):161–8. <https://doi.org/10.1302/2046-3758.35.2000280>.
38. Nissen T, Wynn R. The clinical case report: a review of its merits and limitations. *BMC Res Notes*. 2014;7:264. <https://doi.org/10.1186/1756-0500-7-264>.
39. Parkinson B, Robb C, Thomas M, Thompson P, Spalding T. Factors that predict failure in anatomic single-bundle anterior cruciate ligament reconstruction. *Am J Sports Med*. 2017;45(7):1529–36. <https://doi.org/10.1177/0363546517691961>.
40. The periodic health examination. Canadian Task Force on the Periodic Health Examination. *Can Med Assoc J*. 1979;121(9):1193–254.
41. Röhrig B, du Prel JB, Blettner M. Study design in medical research: part 2 of a series on the evaluation of scientific publications. *Dtsch Arztebl Int*. 2009;106(11):184–9. <https://doi.org/10.3238/arztebl.2009.0184>.
42. Ryan J, Magnussen RA, Cox CL, Hurbank JG, Flanigan DC, Kaeding CC. ACL reconstruction: do outcomes differ by sex? A systematic review. *J Bone Joint Surg Am*. 2014;96(6):507–12. <https://doi.org/10.2106/jbjs.m.00299>.
43. Schardt C, Adams MB, Owens T, Keitz S, Fontelo P. Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Med Inform Decis Mak*. 2007;7:16. <https://doi.org/10.1186/1472-6947-7-16>.
44. Sheiner LB. Is intent-to-treat analysis always (ever) enough? *Br J Clin Pharmacol*. 2002;54(2):203–11.
45. Shore BJ, Nasreddine AY, Kocher MS. Overcoming the funding challenge: the cost of randomized controlled trials in the next decade. *J Bone Joint Surg Am*. 2012;94(Suppl 1):101–6. <https://doi.org/10.2106/jbjs.l.00193>.
46. Weinstein JN, Tosteson TD, Lurie JD, et al. Surgical vs nonoperative treatment for lumbar disk herniation: the spine patient outcomes research trial (sport): a randomized trial. *JAMA*. 2006;296(20):2441–50. <https://doi.org/10.1001/jama.296.20.2441>.
47. Wenger DR. Limitations of evidence-based medicine: the role of experience and expert opinion. *J Pediatr Orthop*. 2012;32(Suppl 2):S187–92. <https://doi.org/10.1097/BPO.0b013e318259f2ed>.
48. Wright JG, Swiontkowski MF, Heckman JD. Introducing levels of evidence to the journal. *J Bone Joint Surg Am*. 2003;85-a(1):1–3.
49. Zantop T, Kubo S, Petersen W, Musahl V, Fu FH. Current techniques in anatomic anterior cruciate ligament reconstruction. *Arthroscopy*. 2007;23(9):938–47. <https://doi.org/10.1016/j.arthro.2007.04.009>.



Bias and Confounding

3

Naomi Roselaar, Magaly Iñiguez Cuadra,
and Stephen Lyman

3.1 Introduction: What Is Bias?

Bias in clinical research refers to any study limitation that precludes the *unprejudiced* consideration of a question [20]. Bias can be considered the biggest enemy of research, as it can distort findings, weaken true associations, or produce spurious associations, destroying the validity of a study. All bias can be categorized as either random or systematic [22]. Random bias occurs when the lack of precision is distributed randomly across the study. Systematic bias occurs when the lack of precision is concentrated among specific subsets of the study population (Fig. 3.1).

Bias can be introduced during many stages of research including design of the study, sample selection, data collection, data analysis, and interpretation of research results [4]. Bias can even occur during publication. That manuscripts reporting statistically significant results are published more often than those that report negative results, results in systematic bias [27]. Each phase of research is susceptible to common types of bias and awareness of these tendencies can help researchers avoid them. Below, bias as it may arise during each phase of research is addressed.

N. Roselaar · S. Lyman (✉)
Hospital for Special Surgery, New York, NY, USA
e-mail: LymanS@HSS.edu

M. I. Cuadra
Cirugía de Rodilla, Traumatología CLC,
Santiago, Chile

3.2 Biases in the Study Design Phase

Fact Box 3.1

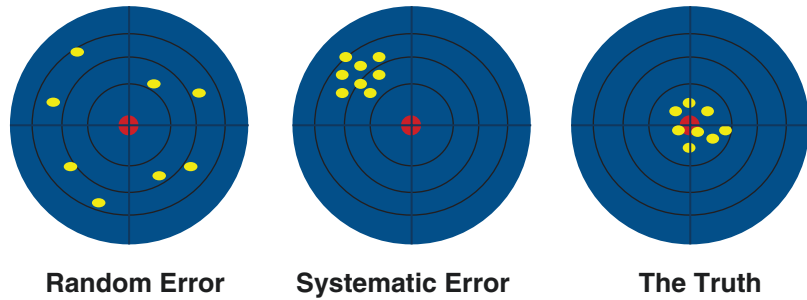
Conceptual bias and confounding are two of the most common types of bias that occur during the design phase of a clinical research study.

Two of the most common types of bias that occur during the design phase of a study are conceptual bias and confounding.

3.2.1 Conceptual Bias

In orthopedics, conceptual bias can occur during the identification and selection of procedures. When selecting procedures for a study, it is important to consider evidence in the literature, specifically support based on randomized controlled trials. Bias can be introduced when procedures are selected on criteria based on personal perception rather than support from the literature. This is particularly frequent for orthopedic procedures as discussed by Lim et al. [17]. Different types of orthopedic procedures were investigated with respect to their degree of RCT evidence and support. Of the total orthopedic procedures assessed, only 37% were supported by at least

Fig. 3.1 A visual representation of how bias can lead to random or systematic error, diverging from the truth



one RCT according to criteria used by the authors [17]. Therefore, the selection of procedures when designing a study sometimes does not occur based on evidence from the literature.

3.2.2 Confounding Bias

The basis for confounding is the presence of one or more risk factors independently related to both the exposure and outcome of a study [24]. Known as confounding variables, confounding factors, or confounders, these variables primarily affect non-randomized observational studies. When unseen, these risk factors can affect the association between the exposure and outcome in a study. Randomized controlled trials are affected by confounding factors to a lesser degree, because the confounding effects are greatly reduced or eliminated by the process of randomization. This is one of the most important reasons why randomized controlled trials are considered the gold standard in clinical research design [30].

3.2.2.1 Measuring and Controlling Confounding Factors

It is impossible to completely eliminate the effects of all confounding factors. However, steps can be taken to reduce the confounding effects of certain variables. If the confounding factor in question is measurable, the effects can be decreased or eliminated [23]. Mitigating the effects of confounding factors is possible only to the degree of precision that the factor is measurable [23]. Confounding bias can be measured using methods such as restriction and matching and can be controlled for using multivariate

analysis [10]. Both restriction and matching methods can be used to mitigate confounding factors at the research design stage [4].

3.2.2.2 Restriction

The principle of restriction involves limiting the ability of a variable to vary in the study population [10]. As Gerhard et al. explain, “A variable has to be able to vary to be associated with an exposure.” In a study of only non-smokers, smoking cannot act as a confounding variable.

Clinical Vignette

A basic orthopedic example of restriction used to negate a measurable confounding variable may be found in the study of ACL injuries in young athletes. Sports may be a confounding factor when considering the association between sex and suffering from an ACL injury in young athletes. Types of sports are associated with both the exposure (sex) and with the outcome (ACL injury frequency). Some sports such as American football or rhythmic gymnastics are either predominantly male or female, while cutting and pivoting sports like soccer and lacrosse have the highest rates of ACL injury. Because the type of sport that an athlete plays can be precisely determined (measured), the study design can be modified to eliminate this confounding effect. Limiting the study to include only participants who play a specific sport (or sports) through the use of specific inclusion and exclusion criteria is one way to control this confounding variable.

3.2.2.3 Matching

Matching is another way to prevent confounding variables from affecting a study. Like restriction, matching is implemented during the design phase of the study. However, instead of restricting the study to certain groups, matching involves creating comparison groups with equal distributions of the specified confounders [10]. The goal of matching is to create two groups that are similar in every way except for the exposure of interest [6]. Note that matching becomes more difficult the more variables you attempt to match on, so it's easier to create a matched comparison group with a smaller number of confounding variables [10]. The trade off, of course, is that your study will be more prone to residual confounding.

Clinical Vignette

There are many cases in the orthopedic literature of matched-controlled studies. One such example is the article, “Total Knee Replacement in Morbidly Obese Patients” from *The Bone & Joint Journal* [1]. In the 4-year assessment of outcomes following total knee replacements by Amin et al., body mass index (BMI) was the exposure of interest. The morbidly obese cohort underwent 41 primary total knee replacements (TKRs). To create a comparison group, the authors selected a group of 41 TKRs in nonobese patients. In order to limit confounding effects of age, sex, diagnosis, type of prosthesis, laterality, and pre-operative functional scores, the control group was matched to the morbidly obese group for all of these variables in order to isolate the effect of morbid obesity on TKR outcomes.

3.2.2.4 Avoiding Bias Due to Confounding Factors

When designing a study, bias due to confounding factors can be mitigated by controlling how patients will be selected and by collecting the appropriate data. For patient selection, strategies

include randomization, restriction, or matching. Randomized trials are considered to have the highest level of evidence because randomization increases the likelihood that *unknown* confounders—i.e., ones that cannot be anticipated—are avoided. When randomization is not possible, restriction and matching (discussed above) are useful. By making sure that the study design includes ways to collect data on potential confounders will also help mitigate their effect during the analysis phase.

3.3 Biases in the Patient Recruitment Phase

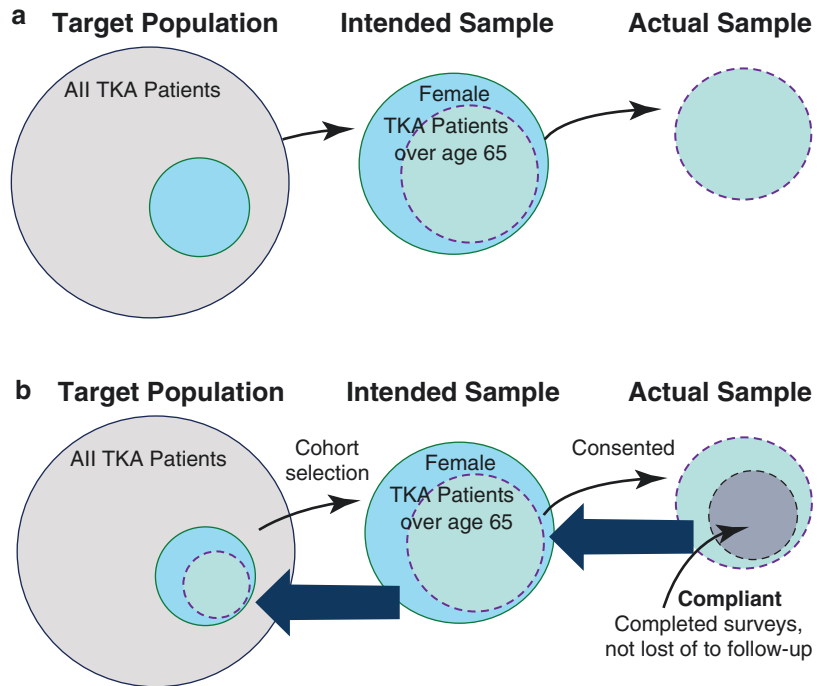
Selection and/or retention bias can occur during the recruitment phase of a study. Selection bias occurs when recruitment methods introduce disproportionate differences between treatment and control (or comparison) groups. Biased selection compromises the internal validity of a study because the actual sample population will be significantly different from the intended population (Fig. 3.2a).

Retention bias—also known as transfer bias—occurs when patients from either the treatment or control (or comparison) group become more likely to leave or be lost from the study, based on some aspect of the study. This type of bias can compromise generalizability, as the study will only be representative of the patients who stayed in the study (Fig. 3.2b).

3.3.1 Selection Bias

Selection bias, which can be random or systematic, occurs on the basis of how subjects are recruited into a study. It is obviously a risk in retrospective studies, where the interest in a known outcome drives the study question. However, selection bias can also occur in prospective studies. In a prospective study, the individuals in comparison groups may differ systematically [16]. For example, a study comparing outcomes of surgical or nonsurgical treatment of anterior cruciate ligament (ACL) injury may reflect dif-

Fig. 3.2 (a) Bias during selection may result in a sample population that is different from the intended study population. (b) Retention bias may result in a sample population that is not generalizable to the intended target population



ferences in the patients who did or did not select surgery—such as level of activity, severity of injury, or age—which affect their outcome as much or more than the treatment itself. Some of these confounding factors could be controlled by collecting relevant data and accounting for them in the analysis. However, other factors that may be systematically associated with either the treatment or control group, such as motivation to achieve pre-injury level of activity, may be difficult or impossible to detect or measure. Still other factors that result in patients choosing surgical versus nonsurgical treatment, and that affect outcome, may be unknown.

Regardless of the study design, choosing the correct control or comparison group is critical for mitigating selection bias. Whenever possible historical controls should be avoided, as they always introduce some bias due to trends over time.

While more common in observational studies, selection bias can also occur in randomized trials. For example, if recruitment occurs through a tertiary care center—as is often the case for RCTs—patients who are recruited may have different disease or socioeconomic characteristics, which

make their outcomes systematically different from patients who would seek care at a community hospital.

Consenting to randomization can be a significant hurdle to achieving unbiased selection in RCTs in orthopedics (and other surgical specialties). Not surprisingly, few patients are willing to consent to be randomized to receive either surgery or no surgery or even a specific type of surgery. For example, in the Spine Patient Outcomes Research Trial (SPORT) [28], a much lower proportion of eligible patients enrolled in the randomized surgical/nonsurgical arm of the study compared with the parallel observational cohort (Fig. 3.3). That the proportion of patients willing to be randomized is small suggests that this group may represent individuals with a specific personality, disease, or other characteristics that may affect outcome in a way that is different from other patients, so that their findings are not generalizable to the general population of patients with back pain. Furthermore, the low number of patients in the trial diminishes the likelihood that randomization can be effective in mitigating the effect of unknown confounders.

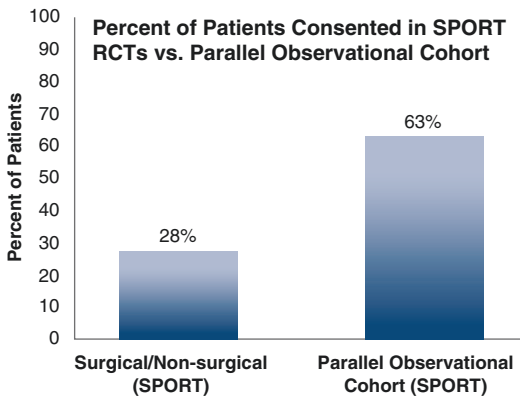


Fig. 3.3 The proportion of eligible patients in the Spine Patient Outcomes Research Trial (SPORT) study consenting to be randomized to surgical versus nonsurgical treatment was much lower than those consenting to enter a parallel observational arm of the study, where patients could choose whether or not to undergo surgery

Narrowing eligibility criteria to limit heterogeneity and reduce confounding and increasing the recruitment period and number of study sites are ways to decrease selection bias in RCTs. Contrary to dogma, in orthopedics and other surgical specialties, where low patient enrollment obviates the benefits of randomization, appropriately controlled, prospective observational studies may ultimately represent a higher level of evidence than RCTs.

3.3.2 Retention Bias

Retention bias occurs when patients who drop out of a study differ from those not lost to follow-up. Examples include patients who die or leave the study due to poor health or because they are unsatisfied with the treatment. Patients may be lost due to non-compliance. A disproportionate loss of patients in one group or another is particularly problematic, as that may introduce error in the analysis of the outcome. For example, if patients leave a study investigating the outcome of a surgical procedure because they undergo revision surgery at an outside hospital, this would cause an underestimation of the rate of complications. Documenting the reasons that patients leave a study can mitigate this type of bias.

Even loss of patients in both groups relatively equally will compromise the generalizability of a study, as the loss may cause the study population to become significantly different from the target population. For example, the specific loss of patients of a particular race or socioeconomic status from a trial due to the distance of a study site from their home would skew the results even if control and treatment groups were affected. Guidelines for the follow-up rates required for publication of randomized controlled trials differ by publication. The *Journal of Bone and Joint Surgery* requires a level 1 randomized controlled trial to have a follow-up rate of at least 80% [14].

In the SPORT study, non-compliance and crossover resulted in a high proportion of retention bias, which undermined the expectations of the study (Fig. 3.4).

Mitigating retention bias may require implementation of measures to reduce loss, such as making follow-up easier by conducting studies at sites accessible to all patients or using outcomes measures that are convenient—such as short form surveys and surveys translated into the patient’s primary language. Limiting crossover may be required. Finally, documenting the reasons for loss of patients from the study may be helpful in accounting for drop out when interpreting the data.

3.4 Informational Biases in the Data Collection Phase

Informational biases can stem from detection or measurement bias by the observer or due to response bias by the study subjects.

3.4.1 Detection or Measurement Bias

Detection bias can occur during the data collection phase of a study when outcomes assessment tools are inappropriate or used inappropriately [27]. Detection bias can arise from poor calibration of a measurement instrument, prejudice in the interviewer or grader, as well as misclassification of exposure or outcome. In orthopedics

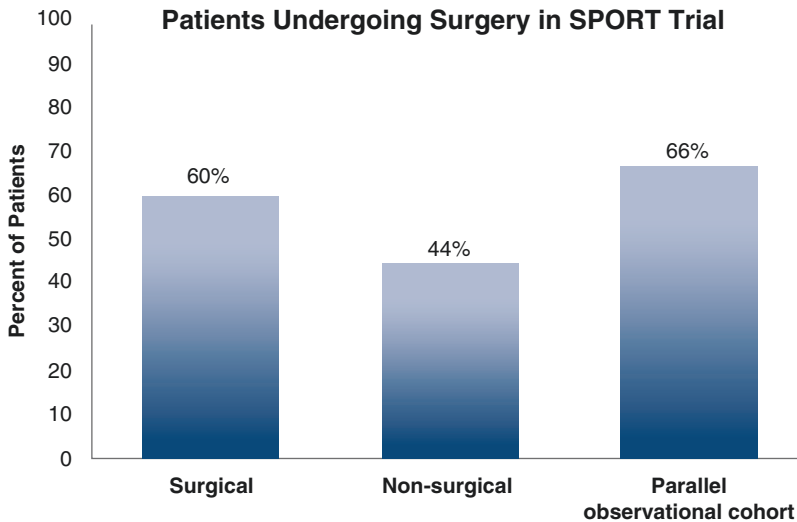


Fig. 3.4 The similarity in the rate of surgery in the surgical and nonsurgical groups made differentiating the outcomes in the two groups impossible in the SPORT study. The proportion of patients in the surgical group who actually

underwent surgery was unexpectedly low because of non-compliance. At the same time, the proportion of patients who started out in the nonsurgical group, but who crossed over and underwent surgery, was unexpectedly high

research, a common cause of detection bias is the failure to use a standardized method to diagnose the condition being studied. This type of bias is more common in retrospective studies, where researchers do not control which diagnostic tools are used for each patient in the study and may not have information on what data or process led to a specific diagnosis [27]. In prospective studies, detection bias can be mitigated by standardizing data collection protocols, including, but not limited to, protocols for how diagnoses are made.

In a study on the association of patellofemoral pain (PFP) and patellofemoral cartilage composition, researchers in the Netherlands compared patients with PFP to healthy controls [26]. All subjects included in the study were diagnosed with PFP based on the presence of at least three criteria from a standardized list. Cartilage composition was assessed in all patients by a single validated magnetic resonance imaging (MRI) method. In this case, researchers minimized the risk of detection bias by using a prospective study design. This allowed them to specify the diagnostic criteria for subject inclusion in the study and standardize the MRI protocol for assessment of cartilage composition in both patients with PFP and the healthy controls.

Blinding can be used to prevent interviewer or grader prejudice from introducing measurement bias. Using clear definitions, validating findings with multiple resources and standardizing measurement protocols are other methods to avoid biases that stem from observer errors.

3.4.2 Response Bias

Subjects in a study may inadvertently give biased responses. The Hawthorne effect refers to a phenomenon where people tend to show bigger improvements, simply on the basis of being observed [29]. Another example of response bias is the tendency for respondents to please the interviewer, especially if the interviewer is the doctor or someone associated with their doctor. Studies asking for responses on sensitive topics can also result in response bias, if patients are too embarrassed to tell the truth.

During the data collection phase of retrospective observational study designs such as case series, case-control studies, and retrospective cohort studies, recall bias can be particularly problematic [30]. Recall bias occurs when surveys or interviews require subjects to recall their

own history, which they may incorrectly remember and report, inadvertently introducing error into the study [25]. Prospective cohort studies avoid recall bias because subject-obtained data is collected in real time [12].

Gabbe et al. [9] compared retrospective, self-reported injury histories of 70 community-level Australian football players with prospective injury surveillance records for the same 12-month period and found a significant difference, to investigate the magnitude of recall bias in a typical type of question that might be included in an orthopedic study. They reported that recall bias increased with the level of detail that was requested [9].

The proportion of community football players who were accurately able to recall their injuries in the past 12 months declined as the level of detail, which was asked about the injury history, increased. All respondents were accurately able to recall the presence or absence of any injury, but fewer than 80% were able to recall the number of injuries or body region where the injury occurred correctly. Accuracy decreased further when respondents were asked to recall number of injuries, where on the body the injury occurred and diagnosis.

The study of Gabbe et al. [9] not only showed how much recall bias could affect research findings but suggested that asking only simple questions could increase the accuracy of responses, when researchers need to ask respondents for their medical history.

Other strategies for mitigating response bias include asking questions related to sensitive topics later in a questionnaire or interview, using validated survey instruments, and seeking responses as early as possible after the exposure or treatment being studied.

3.5 Biases in the Analysis and Data Interpretation Phase

The data analysis and interpretation phase offer opportunities to avoid confounding and other types of bias, which may inadvertently or unavoidably have been introduced in earlier phases. Stratification/subgroup analysis can be

used to reveal confounding. Multivariable modeling can be used to adjust for known potential confounders. Time-to-event analyses can help mitigate bias due to loss to follow-up.

Propensity score matching can be effective in mitigating selection bias.

3.5.1 Propensity Scores

Propensity scores are generated from logistic regression models including all potentially important information that could influence a clinician's treatment decision. Therefore, the model is estimating the likelihood of a patient going down a treatment pathway, rather than an outcome. The coefficients are used to assign each patient a propensity score ranging from 0 to 1 representing the patient's likelihood of receiving a reference treatment (usually the standard of care). A score of 0 indicates no chance of receiving treatment A, while a score of 1 represents absolute certainty of receiving treatment A. A score of 0.5 would represent an unbiased random allocation.

Once propensity scores are assigned to each study subject, they can be used in a variety of ways. For example, a propensity score "match" could be performed, which simulates randomization. A patient with a propensity score of 0.576 for a specific surgery would be matched to a patient from the other treatment type with a propensity score nearest to 0.576. A number of methods exist to optimize this matching process, but when appropriately performed, relevant characteristics of the two groups will be balanced, minimizing potential selection bias. This can be visualized by plotting the standardized differences for each variable before and after matching (Fig. 3.5). A standardized difference is used to compare balance between treatment groups for each independent variable in the dataset. It was proposed by Cohen [3] as Cohen's effect size index for the comparison of two sample mean values [8, 18]. A standardized difference of >0.1 is considered to represent an imbalance in the variable. As can be seen from the example, 14 variables were unbalanced before the match and all variables were balanced after the match. In fact, for all but two of the variables, the balance improved.

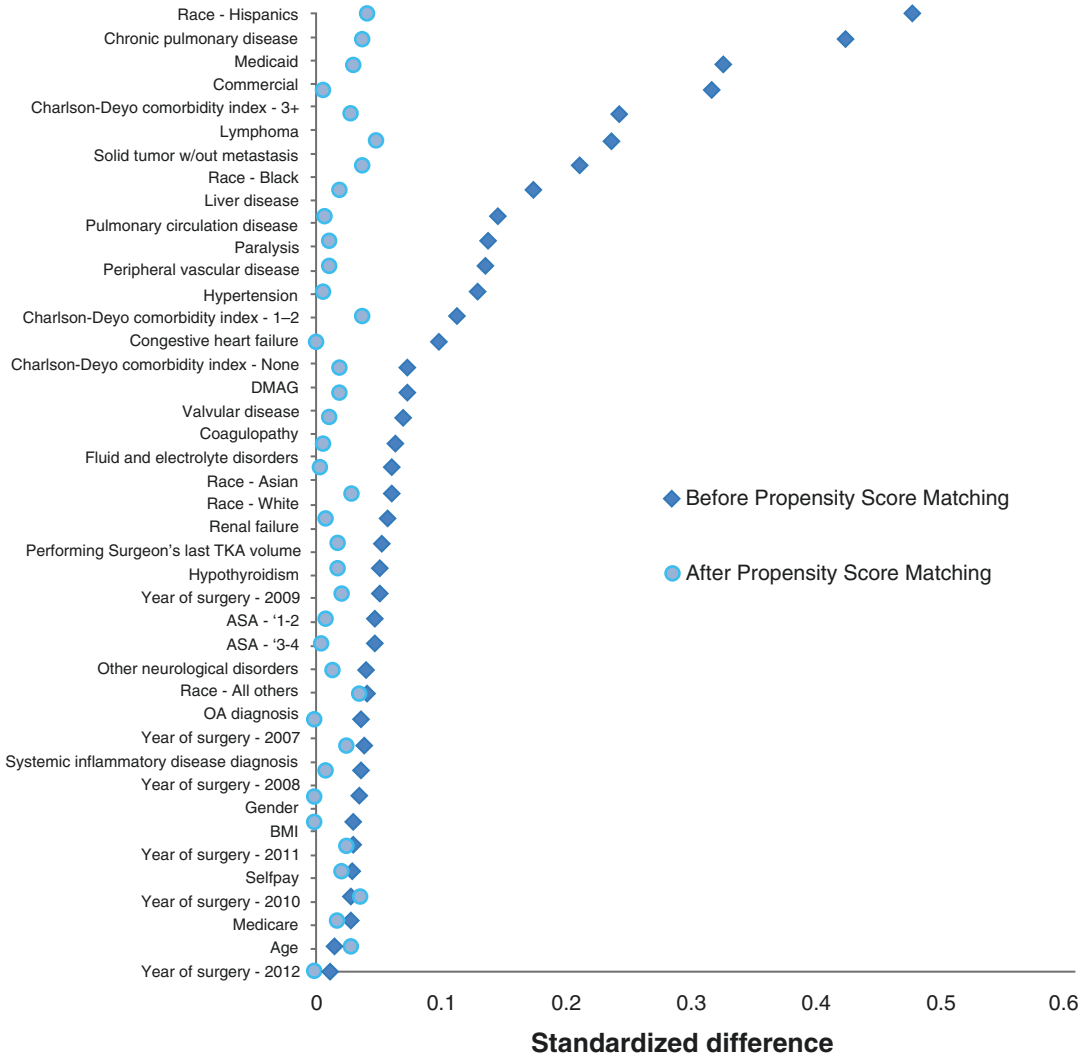


Fig. 3.5 Standardized differences were calculated for the 42 variables (along x-axis) for which patients in a sample cohort undergoing two different knee procedures were matched. The standardized difference (also known as Cohen’s effect size index, along the y-axis) compares two sample mean values in units of the pooled standard deviation so that a standardized difference of 0.1 or greater

denotes meaningful imbalance in the variable. In this example, there is imbalance (standardized difference > 0.1) in 15 variables among patients in the two treatment groups before propensity score matching (dark blue diamonds). However, after propensity score matching (light blue circles), the likelihood these variables affected each patient’s assignment to a particular treatment became more balanced

Whereas randomization eliminates bias, even from unknown confounders, in a strict sense, propensity matching can only balance for known factors. However, because the matching process balances all measured variables, the effects of unmeasured confounders are often also minimized. Despite this, it is incumbent on orthopedic researchers to review the existing literature to

determine a standardized list of variables that best represents the most important information for creating a propensity score. Propensity score matching has been used in orthopedics including arthroplasty and trauma [5, 15].

A drawback of propensity score matching is that a very large number of patients may be needed, especially in the standard of care group

[11]. Moreover, matching frequently omits a large proportion of the available study population when comparison groups are being created. Therefore, inverse probability of treatment weighting (IPTW) is proposed as an alternative to matching to adjust for confounders [2, 13, 21]. With IPTW, the propensity score is used as weights in a weighted regression. Weights restore balance in the clinical characteristics of the two treatment groups and allow for the use of the entire sample, rather than the subset of matched patients.

3.6 Publication Bias

Despite diligent attention to study design and execution, bias can be introduced at the stage of publication. Reviewers are also human and studies have shown that prejudice can affect which studies are published. Emerson et al. (JAMA 2010) used fabricated manuscripts to show that those with positive outcomes were more likely to be accepted for publication than those with negative ones. This article also reported the presence of positive outcome bias in the *Journal of Bone and Joint Surgery (JBJS)* and *Clinical Orthopaedics and Related Research (CORR)* [7]. Another type of bias was reported by Okike et al. (JBJS 2008), who found that submissions to JBJS were more likely to be accepted if they were from the USA or Canada [19]. These studies indicate that readers should be aware of bias, even when reading peer-reviewed papers.

Take-Home Message

- When performing orthopedic research, the effects of confounding factors and bias must be acknowledged.
- Techniques such as randomization, matching, and choice of study design may help mitigate the effects of confounding and bias.

References

1. Amin AK, Clayton RAE, Patton JT, Gaston M, Cook RE, Brenkel IJ. Total knee replacement in morbidly obese patients: results of a prospective, matched study. *J Bone Joint Surg [Br]*. 2006;88-B(10):1321–6.
2. Cassel C, Sarndal C, Wretman J. Some uses of statistical models in connection with the nonresponse problem. In: Madow W, Olkin I, editors. *Incomplete data in sample surveys*. Symposium on incomplete data, proceedings, vol. 3. New York, NY: Academic; 1983.
3. Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. New York: Routledge.
4. Delgado-Rodríguez M, Llorca J. Bias. *J Epidemiol Community Health*. 2004;58:635–41.
5. Duchman K, Gao Y, Pugely A, Martin C, Callaghan J. Differences in short-term complications between unicompartmental and total knee arthroplasty: a propensity score matched analysis. *J Bone Joint Surg Am*. 2014;96(16):1387–94.
6. Dunn WR, Lyman S, Marx R. ISAKOS scientific committee report research methodology. *Arthroscopy*. 2003;19(8):870–3.
7. Emerson GB, Warne WJ, Wolf FM, Heckman JD, Brand RA, Leopold SS. Testing for the presence of positive-outcome bias in peer review: a randomized controlled trial. *Arch Intern Med*. 2010;170(21):1934–9.
8. Flury B, Riedwyl H. Standard distance in univariate and multivariate analysis. *Am Stat*. 1986;40:249–51.
9. Gabbe BJ, Finch CF, Bennell KL, Wajswelner H. How valid is a self reported 12 month sports injury history? *Br J Sports Med*. 2003;37(6):545–7.
10. Gerhard T. Bias: considerations for research practice. *Am J Health Pharm*. 2008;65:2159–68.
11. Guo S, Fraser M. *Propensity score analysis: statistical methods and applications*. 2nd ed. Thousand Oaks, CA: Sage; 2015.
12. Hennekens C, Buring J. *Epidemiology in medicine*. 1st ed. Boston: Little, Brown and Company; 1987.
13. Hirano K, Imbens G. Estimation of causal effects using propensity score weighting: an application to data on right heart catheterization. *Health Serv Outcomes Res Methodol*. 2001;2:259–78.
14. JBJS. JBJS Inc. Journals Level of Evidence [Internet]. *J Bone Joint Surg*. 2015. <https://journals.lww.com/jbjsjournal/Pages/Journals-Level-of-Evidence.aspx>.
15. Jenkinson R, Kiss A, Johnson S, Stephen D, Kreder H. Delayed wound closure increases deep-infection rate associated with lower-grade open fractures: a propensity-matched cohort study. *J Bone Joint Surg Am*. 2014;96(5):380–6.
16. Jennings JM, Sibinga E. Understanding and identifying bias in research studies. *Pediatr Rev*. 2010;31(4):161–2.
17. Lim HC, Adie S, Naylor JM, Harris IA. Randomised trial support for orthopaedic surgical procedures. *PLoS One*. 2014;9(6):e96745.
18. Normand S, Landrum M, Guadagnoli E, Ayanian J, Ryan T, Cleary P, et al. Validating recommendations for coronary angiography following an acute myocardial infarction in the elderly: a matched analysis using propensity scores. *J Clin Epidemiol*. 2001;54:387–98.
19. Okike K, Kocher MS, Mehlman CT, Heckman JD, Bhandari M. Publication bias in orthopaedic

- research: an analysis of scientific factors associated with publication in *The Journal of Bone and Joint Surgery (American Volume)*. *J Bone Joint Surg Am*. 2008;90(3):595–601.
20. Pannucci CJ, Wilkins EG. Identifying and avoiding bias in research. *Plast Reconstr Surg*. 2010;126(2):619–25.
 21. Rosenbaum P. Model-based direct adjustment. *J Am Stat Assoc*. 1987;82:387–94.
 22. Shapiro S. Causation, bias and confounding: a hitchhiker's guide to the epidemiological galaxy Part 1. Principles of causality in epidemiological research: time order, specification of the study base and specificity. *J Fam Plann Reprod Health Care*. 2008a;34(2):83–7.
 23. Shapiro S. Causation, bias and confounding: a hitchhiker's guide to the epidemiological galaxy: Part 2. Principles of causality in epidemiological research: confounding, effect modification and strength of association. *J Fam Plann Reprod Health Care*. 2008b;34(3):185–90.
 24. Shapiro S. Causation, bias and confounding: a hitchhiker's guide to the epidemiological galaxy Part 3: Principles of causality in epidemiological research: statistical stability, dose- and duration-response effects, internal and external consistency, analogy and biological plausibility. *J Fam Plann Reprod Health Care*. 2008c;34(4):261–4.
 25. Smith J, Noble H. Bias in research. *Evid Based Nurs*. 2014;17(4):100–1.
 26. van der Heijden RA, Oei EHG, Bron EE, van Tiel J, van Veldhoven PLJ, Klein S, et al. No difference on quantitative magnetic resonance imaging in patellofemoral cartilage composition between patients with patellofemoral pain and healthy controls. *Am J Sports Med*. 2016;44(5):1172–8.
 27. Viswanathan M, Berkman ND, Dryden DM, Hartling L. Assessing risk of bias and confounding in observational studies of interventions or exposures: further development of the RTI item bank. *Agency Health Res Qual*. 2013:1–22.
 28. Weinstein JN, Tosteson TD, Lurie JD, Tosteson ANA, Hanscom B, Skinner JS, et al. Surgical vs nonoperative treatment for lumbar disk herniation the Spine Patient Outcomes Research Trial (SPORT): a randomized trial. *JAMA*. 2006;296(20):2441–50.
 29. Wickstrom G, Bendix T. The “Hawthorne effect”—what did the original Hawthorne studies actually show? *Scand J Work Environ Health*. 2000;26(4):363–7.
 30. Zlowodzki M, Jönsson A, Bhandari M. Common pitfalls in the conduct of clinical research. *Med Princ Pract*. 2006;15:1–8.



Ethical Consideration in Orthopedic Research

4

Jason L. Koh and Diego Villacis

4.1 Introduction

For thousands of years, science has sought answers to questions regarding health and the human body. From the start, there has been a dilemma with defining the boundary of ethical research while still pushing to advance scientific knowledge for the benefit of humanity. There came to be a universal understanding that protection of the human participants must take priority. Unfortunately, our society has paid heavy costs in the past when failing to ensure ethical research. From these early beginnings, principles arose to guide the conduct and review of human subject research. These principles were formalized by the development of regulations and governing bodies to review the creation of research protocols.

To begin one must understand the purpose of health research. A study should be designed to develop new or confirm knowledge that promotes

health, prevents disease and injuries, and improves the diagnosis and treatment of disease and injuries [1]. In order to justify the use of human subjects, the potential risks to the research participant must be reasonable in relation to the potential benefits to the participant or future patients and be sensible because of the importance of the knowledge which might be gained [1]. For any research study, the full benefits and risk can never truly be known ahead of time, nor can the effects of the study be determined until a study is completed. This quandary is a reality for researchers that highlights the importance of oversight and regulatory bodies to ensure that human research studies stay true to guiding ethical principles.

4.2 History

The contemporary viewpoint with emphasis on the individual and the importance to protect the research subject was initiated long before the development of modern research ethics. Hippocrates (460 BC) is regarded by many as the best known of early scientists to promote physician constraint. His beliefs are well quoted in the *Hippocratic Oath*, which requires physicians to practice medicine in such a way that ultimately benefits the patient while avoiding mischievous behavior or behavior that is not in the best interest of the patient, “Do no harm” [14, 25]. However,

J. L. Koh · D. Villacis (✉)

Department of Orthopaedic Surgery, NorthShore
University Health System, Evanston, IL, USA

Northshore Orthopedic Institute, Evanston, IL, USA

Pritzker School of Medicine, University of Chicago,
Chicago, IL, USA

e-mail: Jkoh@northshore.org;

dvillacis@northshore.org

for the purpose of this section, we will focus on developments from the modern era, starting with the aftermath of the atrocities suffered during World War II.

The Nuremberg Military Tribunal after World War II shed light on the sadistic and horrifying “research” conducted by Nazi German scientists. Nazi research included forced human exposure to the effects of freezing, incendiary devices, mustard gas, and other weaponized agents. The rulings handed down from the Nuremberg trial served as an outline for the required elements of conducting ethical human research.

Fact Box 4.1

Initially known as rules for “Permissible Medical Experiments,” this became known as the “Nuremberg Code.” The leading ethical principles that emerged were a requirement for voluntary consents, evidence of scientific merit, benefits outweighing the risks, and the ability of research subjects to terminate participation in the study at any-time [23, 27].

Although not formally adopted into law or as any part of a professional ethical code, it served a significant influence for the development of contemporary ethical guidelines.

Fact Box 4.2

Subsequently, the Declaration of Helsinki in 1964 was developed by the World Medical Association (comprised of almost all national medical associations) to provide guiding principles regarding human subject research. Key among these principles are a respect for the individual and their right to make informed choices regarding participation in research [9]. The individual’s interests are placed above those of “science and society.”

Multiple revisions have been made over the years; however, the principles set forth by the “Nuremberg Code” remain the foundation for the conduct of ethical research all over the world.

During early times of modern human research ethics, audible outcry was forming in the United States. Ethical concern was mounting regarding the 1936 publication of a paper funded by the federal government entitled “The Tuskegee Study of Untreated Syphilis in the Negro Male.” The purpose of the study was to investigate the effects of untreated syphilis starting in 1932, essentially tracking the natural history. The study would remain in progress until 1972. What would make the study infamous was the continued withholding of treatment even after the acceptance of penicillin as the standard of care in 1945. Subjects were led to believe that they were being treated when in fact nothing was being done. It took nearly 40 years from the initial publication of the study until stories in the press created public outcry on a national level, and the study was concluded [13].

The subsequent attention gained by the shocking truths of the Tuskegee study leads to congressional hearings to address the matter of ethical conduct in human subject research. As a direct response to these atrocities and those of other similar situations of ethical abuse, the National Research Act of 1974 was passed to address ethical concerns [7]. The congress called for the establishment of “The National Commission for the Protection of Human Subjects and Biomedical and Behavioral Research” and the establishment of institutional review boards (IRBs). The “National Commission” was tasked with identifying the basic ethical principles underlying human subject research and develops guidelines for ensuring ethical principles are followed [3]. The commission met between 1975 and 1978 releasing a series of reports with their last report, a summary of their deliberations regarding ethical principles, released in 1979. This final report was entitled “Ethical Principles and Guidelines for the Protection of Human Subjects of Research” and would come to be known as the Belmont Report.

Fact Box 4.3

The Belmont Report established boundaries between clinical practice and unnecessary research, basic ethical principles, and fundamental applications of ethical research [11].

The three basic ethical principles identified by the Belmont Report include “respect for persons” which asserts the importance of respecting autonomy, “beneficence” which requires actions that do not cause harm and aim to maximize benefit while minimizing harm, and “justice” which requires a shared burden of benefits and risk when choosing research subjects so as not to exploit a population. These principles set forth by the Belmont Report share importance and are to be utilized as a “framework” when designing human subject research protocols [8]. To ensure that research does meet the abovementioned criteria, IRBs were created (Table 4.1).

4.3 Institutional Review Board (IRB)

The US Department of Health and Human Services (HHS) under the guidance of the Belmont Report findings set up requirements that form the basis for institutional review boards. The goal of an IRB is to regulate human subject research by advocating, upholding, and maintaining the rights of research participants [11]. IRBs have become widespread and are engaged in all HHS- and National Institutes of Health (NIH)-funded studies. The initial requirements proposed by HHS were formalized in 1991 by the US Federal Policy for the Protection of Human Subjects. Protection of Human Subjects became better known as “Common Rule,” which provided specific direction for the structure and organization of IRBs, the requirements for obtaining informed consent, and the requirement that institutions provide written assurance that

Table 4.1 Nuremberg Code 1947 [27]

1.	The voluntary consent of the human subject is absolutely essential
2.	The experiment should be such as to yield fruitful results for the good of society, unprocurable by other methods or means of study, and not random and unnecessary in nature
3.	The experiments should be so designed and based on the results of animal experimentation and knowledge of the natural history of the disease or other problem under study that the anticipated results will justify the performance of the experiment
4.	The experiment should be so conducted as to avoid all unnecessary physical and mental suffering and injury
5.	No experiment should be conducted where there is a prior reason to believe that death or disabling injury will occur, except perhaps, in those experiments where the experimental physicians also serve as subject
6.	The degree of risk to be taken should never exceed that determined by the humanitarian importance of the problem to be solved by the experiment
7.	Proper preparations should be made and adequate facilities provided to protect the experimental subject against even remote possibilities of injury, disability, or death
8.	The experiment should be conducted only by scientifically qualified persons. The highest degree of skill and care should be required through all stages of the experiment of those who conduct or engage in the experiment
9.	During the course of the experiment, the human subject should be at liberty to bring the experiment to an end if he has reached the physical or mental state where continuation of the experiment seems to him to be impossible
10.	“During the course of the experiment, the scientist in charge must be prepared to terminate the experiment at any stage, if he has probable cause to believe, in the exercise of the good faith, superior skill, and careful judgment required of him, that a continuation of the experiment is likely to result in injury, disability, or death to the experimental subject”

they are compliant with federal regulations [26]. Medical and academic institutions had already in large part created their own IRBs to review clinical research projects. This was formalized with “Common Rule” so as to further regulate that IRBs review all research protocols for compliance with ethical guidelines [4].

An important concept for IRB review is understanding when review is necessary. The necessity for review relates directly to determining the risk of a proposed study. Per the IRB risks are split into three broad categories: less than minimal risk, minimal risk, and greater than minimal risk. A study with less than minimal risk means there is no known physical, psychological, or economic risk to the subject [2]. Such a situation would deem the study exempt from IRB review. However, exempt status must be determined in a formal process with adequate documentation and not simply be determined by the researcher. Relevant to orthopedic research, a study may also be deemed exempt if it limits human subjects to one or more of the following categories: educational practices and assessments, interviews or observations of public behavior, and studies of public data or specimens without accompanying information that might permit subject identification [26]. When a study poses minimal risk, it may qualify for expedited review by select members of the IRB committee. Common examples of such studies include observational studies reviewing medical data that is collected as part of routine medical care, medical chart review with no unique identifiers included in the record, and questionnaire or survey studies where it is unlikely the questions would cause emotional distress in the patient [2]. Studies that are determined to pose “greater than minimal risk” as defined by risk beyond those typically encountered by the patient require a full-committee review by the IRB [2]. Once under review the committee will assess a broad list of factors prior to acceptance including study design, deception or withholding of information, risks and benefits, selection of subjects, identification of research participants, protection of patient privacy, process of obtaining informed consent, informed consent forms, qualifications of investigators, and potential conflict of interests [19]. After acceptance, for studies lasting longer than 1 year, it is required that the study be reviewed annually [19] (Fig. 4.1).

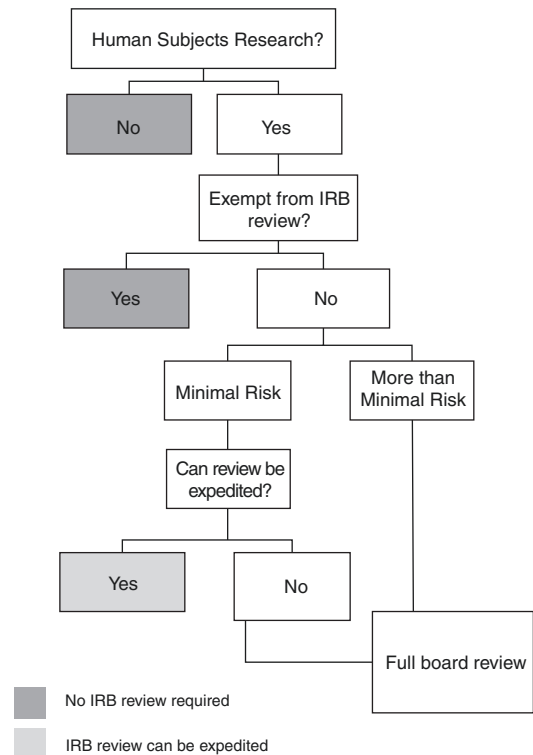


Fig. 4.1 Regulatory questions to ask prior to IRB [11]

4.4 Privacy Regulations: The United States and Europe

Additional regulatory guidelines for human subject research were enacted in the United States with the passing of the Health Insurance Portability and Accountability Act (HIPAA) in 1996.

Fact Box 4.4

The broader goal of the HIPAA was “improve portability and continuity of health insurance coverage in the group of individual markets, to combat waste, fraud, and abuse in health insurance and health care delivery, to promote the use of medical savings accounts, to improve access to long-term care services and coverage, to simplify the administration of health insurance, and for other purposes” [12].

Although it was to have an initial impact on health-care providers and insurance carriers, it would also go on to affect human subject research through its impact on protection of patient privacy. This focus on protection of patient privacy was expanded with a provision in 2003 requiring all “covered entities,” essentially any personnel with access to patient information, to be compliant with the HIPAA privacy rule. The goal of the HIPAA privacy rule is to regulate the use and disclosure of specific “individually identifiable health information,” termed protected health information (PHI) [6]. A patient’s PHI is considered to be information pertaining to their past, present, or future physical or mental health, the provision of health care to an individual, and the past, present, or future payment for the provision of health care to the individual [17]. Similar legislation to HIPAA has been recently enacted in the European Union in 2016 [24]. In order to help ensure the privacy of PHI during human subject research, the IRB is tasked with allowing the transmission of only the minimal amount of information necessary. The IRB must also make sure that this information is also adequately protected when transmitted or stored electronically with tools such as encryption or password protection. In the United States, violations of this act can be punishable

by public disclosure of the breach, limitations on research, fines, and imprisonment (Fig. 4.2).

4.5 Informed Consent

The process of informed consent is a fundamental pillar to ethical human subject research. In essence the participant must be able to both understand and clearly verbalize a study’s purpose, methods, risk, benefits, and alternative to participation in order to be considered “informed.” The participant must be allowed to freely decide whether to participate in the study both at the beginning and throughout the study with the option to discontinue at any time. For situations in which the potential subject is a minor or adult of otherwise limited mental capacity, a proxy can be used with the same requirements of informed consent [22]. Informed consent is designed so as to demonstrate respect for the autonomy of potential and enrolled participants. Therefore, informed consent must be obtained in a voluntary and noncoercive manner. This goal of voluntary enrollment can be challenging when considering potential vulnerability to autonomy in situations where there is a large inequality in authority between researcher and enrolled or potential participant (Fig. 4.3).

Fig. 4.2 List of PHI [24]

1. Names
2. Contact information (e.g., phone or fax #, website or internet protocol [IP] or electronic mail addresses, geographic address smaller than State, except first three digits of zip code)
3. Identifying dates (more detailed than year, e.g., birth, death, admission, discharge)
4. Age over 89 years (unless listed as 90 or older)
5. Social security, medical record, insurance identification number
6. Vehicle identification numbers (e.g., serial numbers, license plate numbers)
7. Device identification or serial numbers
8. Certificate or license numbers
9. Biometric identity (e.g., voice or retinal print, fingerprint, full face image)
10. Any other unique account numbers or material

Fig. 4.3 Paragraph VIII.A. of the American of Orthopaedic Surgeons' *Code of Medical Ethics and Professionalism for Orthopaedic Surgeons* [1]

1. The orthopaedic surgeon must explain to the patient in terms the patient can understand the proposed treatment, its likely effect on the patient, and purpose of the research. Orthopaedic surgeons must provide at least the degree of information that is required by applicable state and federal law, which will include at a minimum information on the purpose of the research, its potential side effects, alternatives and risks of the proposed treatment as well as the method, purpose, conditions of participation and the opportunity to withdraw from the research protocol without penalty.
2. The patient must understand for what they are providing consent. The orthopaedic surgeon must ensure that the patient has understood the basic information and has engaged in rational decision-making in deciding to participate in the research; and
3. The patient's consent must be voluntary. Voluntary consent requires that the patient agreeing to participate in the project has a full understanding of all alternative treatments beyond the research protocol. The orthopaedic surgeon must believe that the patient's consent is free from undue or overbearing influences, e.g., fear of the loss of care or medical benefits if the patient declines to participate.

4.6 Evaluation of Human Research Protocols

Fact Box 4.5

There are seven main principles that can be utilized to guide protocol design for human subject research: social and clinical value, scientific validity, fair subject selection, favorable risk-benefit ratio, independent review, informed consent, and respect for potential subjects [10].

The principles for evaluation of human research protocols are nicely summarized by the NIH with their published recommendations for ethical study design [22]. Informed consent and respect for potential subjects were discussed in the previous section. Below is a brief description of the five remaining principles. Clinical and social value refers to the overriding concern of whether a study will potentially provide answers that lead to a new and significant discovery of information. Also, will this information lead to a tangible impact on patient care both for the individual and society as a whole? When looking for an answer, the study must be

scientifically valid in order to give justification to the study's conclusion. Otherwise the study is a waste of resources and undue risk to the human subjects. During study design the selection of subjects should include concerns for the risk burden by participants and potential for participants to also enjoy benefits from the research study [22]. As discussed throughout this chapter, a heavily weighed factor in any study design is creating a favorable risk-benefit ratio. Since by definition a research study is investigating the unknown, it is impossible to truly know all the risk and benefits of an intervention. However, by aiming to minimize risk and maximize benefits, a study can attempt to avoid foreseeable safety issues or potential harm to the participants. Finally, study protocols must undergo an independent review to ensure the abovementioned principles are upheld. This is the role of an institution's IRB (Fig. 4.4).

1. Social and clinical value
2. Scientific validity
3. Fair subject selection
4. Favorable risk- benefit ratio
5. Independent review
6. Informed consent
7. Respect for potential and enrolled subjects

Fig. 4.4 The seven ethical principles of human research protocol design [10]

4.7 Responsibilities of Principal Investigator

Fact Box 4.6

The principal investigator (PI) of a research study is “responsible for proposing, designing, and reporting the research” [1].

Based on if there is funding associated with the project, the PI is also responsible for the dispersal of funds. A key element brought forth by the American Academy of Orthopaedic Surgeons (AAOS) is that although the PI may have an opinion on the clinical topic being investigated the study can only be justified if there is debate in the medical community regarding the clinical question being pursued. Delegation of portions of the study to others involved in the study is permitted, but “delegation does not relieve the PI of responsibility for work conducted by other individuals” [1]. A frequent issue in the field of research that is infrequently discussed is proper credit and authorship for work done. The PI is responsible for ensuring that any articles relevant to the research study include “appropriate credit for individuals contributing importantly to the research” [1]. Each author of a publication should be able to individually justify the content of the publication regardless of specific contribution (Fig. 4.5) [1].

10. An orthopaedic surgeon shall warrant that he or she has made significant contributions to the conception and design or analysis and interpretation of the data, drafting the manuscript or revising it critically for important intellectual content, and approving the version of the manuscript to be published.”

12. An orthopaedic surgeon shall credit with authorship or acknowledge and not exclude those individuals who substantially contributed to the proposed research, the analysis and interpretation of the data, and the drafting and revising of the final article or report.

Fig. 4.5 Mandatory Standards 10 and 12 from *Standards of Professionalism on Research and Academic Responsibilities*

4.8 Sham Surgery

Utilization and discussion of sham surgery in orthopedic research are limited [20, 30]. The concern is obvious, and there is difficulty in designing a study that can safely minimize risks while presenting a potential benefit to justify these risks. However, we know from robust data that subjective outcomes can significantly be altered by placebo effect [15, 16]. Therefore, a study design that can account for placebo effect would be ideal for determining the true effect on an intervention [16]. In addition, sham surgery allows for the blinding of the subject and the researcher collecting patient data. It is difficult to conceal the scar or incision from both the patient and individual collecting data. Also, there is evidence to support the notion that patients have a bias toward favorable outcome with surgical intervention. Patients want to believe they chose the correct treatment option [28, 29]. This all leads to the conclusion that sham surgery can be acceptable but only if done for an appropriate procedure and the study is designed so as to minimize risk to the human subjects. A common application is when investigating clinical intervention that is strongly suspected of having little benefit to placebo effect. An example of such a trial was conducted by Moseley et al., investigating arthroscopic knee surgery vs. placebo surgery for knee osteoarthritis that had failed 6 months of conservative treatment (Table 4.2) [21].

Table 4.2 Guidelines for sham surgery [20]

1. There is skepticism regarding the therapeutic merits of a particular treatment
2. There are disagreements about the perceived benefits of a particular procedure compared with the placebo
3. Benefits might be due to the “experience of surgery” and the postoperative care regimen
4. Risks are reduced as far as possible in the sham surgery arm without compromising trial design
5. There is a lack of a superior therapy

4.9 Funding and Potential Conflict of Interest (COI)

Traditionally, there was minimal support from the industry for orthopedic research. Research projects came from academic institutions as well as research-active individuals or practices. However, in the past 3 decades, orthopedic surgery has undergone a transformation with the development and flourishing of an “orthopedic industry” [5]. Much of the industry-fueled innovation has come in terms of implant development. During the same period, there has been a shift in research funding from predominantly charitable or government funded to one where industry funding represents a majority [5]. This includes industry support for education, meetings, and conferences. Although increased industry support clearly has the power to discover positive advancements in the treatment of patients, it also produces a potential bias and situation of potential conflict. The potential for conflicts has led to most societies and journals in the orthopedic community requiring full statements of disclosure when presenting any research work. However, successfully navigating the potential conflict of interests in regard to industry funding is still an active challenge.

In order to better understand the arena, it is important to grasp the three parties who have distinct interests in industry-funded research: the

researcher, the research institution, and the corporation funding the research [1]. These interests can vary drastically and do not necessarily run in parallel. Thankfully, many institutions have created structures that assist in negotiating research funding so as to help eliminate or minimize conflicts of interest. As may seem obvious, “ethical problems may arise when the researcher or the research institution have direct financial interest in the research program” [1]. However, for the researcher or research institution, financial bias may not be obvious and can often be subtle. The Academy (AAOS) references two ethical principles for researchers to fall back on when faced with economic conflicts of interest (refer to Table 4.3) [1]. One can conclude from these two principles set forth by AAOS that once someone has received funding from a corporation they should not buy or sell the corporations stocks during the entirety or their involvement in the project. Also, someone who has developed an implant from which they will receive royalties should avoid doing research on that implant, leaving the research to a disinterested third party who has no potential financial interest from use of the device [1]. That does not exclude researchers from being able to serve as consultants for a corporation, given that the compensation is in line with their efforts.

Table 4.3 Ethical principles for economic conflict of interest [1]

1. A researcher ethically may share the economic rewards of his or her efforts. If a drug, device, or other products becomes financially remunerative, the researcher may receive profits that reasonably resulted from his or her contribution. The Academy’s *Standards of Professionalism on Orthopaedic Surgeon-Industry Relationships* and the *Code of Medical Ethics and Professionalism for Orthopaedic Surgeons* explicitly permit an orthopedic surgeon to receive royalties. However, ethically the researcher may not reap profits that are not justified by the value of his or her actual efforts
2. Potential sources of bias in research should be eliminated, particularly where there is a direct relationship between a researcher’s personal interests and potential outcomes of the research

Clinical Vignette

Below are examples of unethical behavior regarding conflict of interests as stated by the AAOS:

- Knowingly negotiating for more funding than is appropriate to support the project and related institutional and departmental overhead costs
- A researcher’s selling or purchasing stock in a company whose orthopedic device is being tested by that orthopedic surgeon-researcher
- A researcher accepting financial incentives to alter data

- A researcher accepting excessive remuneration by the funding corporation for evaluating or interpreting data about that corporation's products
- A failure to disclose research or consulting arrangements with the funding corporation when reporting about research on devices manufactured by that corporation

Clinical Trials Registration

Another aspect of the ethical conduct of research is in reporting results accurately and completely. Concerns regarding selective publication of data (e.g., only positive information in a pharmaceutical trial) that could result in the reporting of misleading conclusions have prompted the International Committee of Medical Journal Editors (ICMJE) to require clinical trials registration of any study that assigns patients to an intervention [18]. For publication to be considered, these studies must be prospectively registered in a publically accessible database, such as clinicaltrials.gov or the World Health Organization (WHO) International Clinical Trials Registry Platform [18].

Take-Home Messages

- The essence of ethics regarding human subject research is respect for the participant's autonomy and well-being.
- The modern era began with the end of World War II and the guidelines put forth with the subsequent Nuremberg Code of 1947.
- The Belmont Report identified three basic ethical principles: respect for persons, beneficence, and justice.
- Institutional review boards were established to regulate human subject research and ensure adherence ethical practices.
- Informed consent requires a patient to be able to understand and verbalize a study's purpose,

methods, risks, benefits, and alternatives to participation.

- There are seven principles to guide human subject research protocol design: social and clinical value, scientific validity, fair subject selection, favorable risk-benefit ratio, independent review, informed consent, and respect for potential subjects.
- Although there is justified concern, there is still a role for sham surgery when a protocol can be designed that minimizes risk while presenting a potential robust benefit.
- There is a growing concern for potential conflict of interest related to industry funding, and therefore it is critical that a researcher be aware of several ethical principles pertinent to conflict of interest when navigating this area.
- Clinical trials registration is important ethically to reduce the risk of selective publication; it has also become a standard for publication in most major medical journals.

References

1. American Academy of Orthopaedic Surgeons, Opinion on Ethics and Professionalism. Ethics in health research in orthopaedic surgery. Adopted October 1994. Revised December 1995, May 2002, July 2003, September 2005, and September 2016.
2. Assessing Level of Risk and Type of IRB Review. Research compliance news. University of South Alabama. 2008. www.southalabama.edu/research-compliance/pdf/compliancencenews0908.pdf. Accessed 1 Feb 2018.
3. Brody B. The ethics of biomedical research. New York: Oxford University Press; 1998.
4. Brown JG. Department of Health and Human Services. Office of Inspector General. Institutional review boards: their role in reviewing approved research. Office of Evaluations and Inspections; 1998.
5. Carr AJ. Which research is to be believed. *J Bone Joint Surg (Br)*. 2005;87-B:1452–3.
6. Clinical Research and the HIPAA Privacy Rule. NIH Publication Number 04-5495. 2004. http://privacyruleandresearch.nih.gov/clin_research.asp. Accessed 29 Jan 2018.
7. Cobb WM. The tuskegee syphilis study. *J Natl Med Assoc*. 1973;65:345–8.
8. Cohen J. History and ethics of human subjects research. Collaborative Institutional Training Initiative. CITI Program. 2017. Accessed 28 Jan 2018.

9. Declaration of Helsinki History Website. Ethical principles for medical research. JAMA Netw. Accessed 26 Feb 2018.
10. Emanuel EJ, Wendler D, Grady C. What makes clinical research ethical? JAMA. 2000;283:2701–11.
11. Friedman EA. Ethical issues in clinical research. In: Supino PG, Borer JS, editors. Principles of research methodology: a guide for clinical investigators. Ed 1 ed. New York: Springer; 2012. p. 233–54.
12. Health Insurance Portability and Accountability Act of 1996. Public Law 104–191. 104th Congress. 1996. <http://www.gpo.gov/fdsys/pkg/PLAW-104publ191/pdf/PLAW-104publ191.pdf>. Accessed 30 Jan 2018.
13. Heller J. Syphilis victims in U.S. study went untreated for 40 years. New York Times (New York). 1972;1:8.
14. The hippocratic oath: today. Doctors' diaries. WGBH Educational Foundation. 1964. <http://www.pbs.org/wgbh/nova/body/hippocratic-oath-today.html>. Accessed 28 Jan 2018.
15. Hrobjartsson A, Gotzsche PC. Is the placebo powerless? An analysis of clinical trials comparing placebo with no treatment. N Engl J Med. 2001;344:1594–602. Erratum in: N Engl J Med;345:304.
16. Hrobjartsson A, Gotzsche PC. Placebo interventions for all clinical conditions. Cochrane Database Syst Rev. 2004;3:CD003974.
17. HHS. Summary of the HIPAA Privacy Rule. <http://www.hhs.gov/ocr/privacy/hipaa/understanding/summary/privacysummary.pdf>. Accessed 30 Jan 2018.
18. International Committee of Clinical Journal Editors. Clinical Trials. <http://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html>. Accessed 27 Feb 2018.
19. Mazur DJ. Evaluating the science and ethics of research on humans: a guide for IRB members. Baltimore: The Johns Hopkins University Press; 2007.
20. Mehta S, Myers TG, Lonner JH, Huffman GR, Sennett BJ. The ethics of sham surgery in clinical orthopaedic research. J Bone Joint Surg Am. 2007;89:1650–3.
21. Moseley JB, O'Malley K, Petersen NJ, Menke TJ, Brody BA, Kuykendall DH, Hollingsworth JC, Ashton CM, Wray NP. A controlled trial of arthroscopic surgery for osteoarthritis of the knee. N Engl J Med. 2002;347:81–8.
22. NIH & Clinical Research. Ethics in Clinical Research. http://clinicalresearch.nih.gov/ethics_guides.html. Accessed 26 Jan 2018.
23. Nuremberg Code [from Trials of War Criminals Before the Nuremberg Military Tribunals Under Control Council Law No. 10. Nuremberg, October 1946–April 1949. Washington, DC: U.S. G.P.O., 1949–1953].
24. Official Journal of the European Union. Legislation L119. <http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AL%3A2016%3A119%3ATOC>. Accessed 27 Feb 2018.
25. Owsei T, Temkin C. Ancient medicine. Selected papers of Ludwig Edelstein Johns. Baltimore: Hopkins University Press; 1987.
26. Public, Protection of Human Subjects, Basic HHS Policy for Protection of Human Research Subjects, Title 45 CFR Part 46, Subpart A. 2005. <http://ohsr.od.nih.gov/guidelines/45cfr46.html>. Accessed 28 Jan 2018.
27. Shuster E. Fifty years later: the significance of the Nuremberg Code. N Engl J Med. 1997;337:1436–40.
28. Sussman MD. Ethical requirements that must be met before the introduction of new procedures. Clin Orthop Relat Res. 2000;378:15–22.
29. Sussman MD. Ethical standards in the treatment of human subjects involved in clinical research. J Pediatr Orthop. 1998;18:701–2.
30. Wolf BR, Buckwalter JA. Randomized surgical trials and “sham” surgery: relevance to modern orthopaedics and minimally invasive surgery. Iowa Orthop J. 2006;26:107–11.

Conflict of Interest

5

Michael Hantes and Apostolos Fyllos

5.1 Introduction

A conflict of interest is a set of circumstances that creates a risk that professional judgment or actions regarding a primary interest will be unduly influenced by a secondary interest.

This is the definition of conflict of interest in medical science given in 2009 by the Institute of Medicine of the National Academies [10]. With fewer government funds available for research and education worldwide, many researchers are turning to industry for research support. Industry relationships with academic biomedical researchers are extensive [3, 21]. In the USA, 71% of research and development funding comes from industry, followed by government (21%) and private foundations (4%) [2]. Despite their benefits, relationships with industry create conflicts of interest that can undermine the primary goals of medical research. Several systematic reviews and other studies provide substantial evidence that clinical trials with industry ties are more likely to have results that favor industry [13, 19]. Other sources of conflict of interest, other than financial, exist as well.

M. Hantes (✉) · A. Fyllos
Department of Orthopaedic Surgery, Faculty of
Medicine, University of Thessalia, University
Hospital of Larissa, Larissa, Greece
e-mail: hantesmi@otenet.gr

Fact Box 5.1

With fewer government funds available for research and education worldwide, many researchers are turning to industry for research support. In the USA, 71% of research and development funding comes from industry, followed by government (21%) and private foundations (4%).

5.2 Definition of Conflict of Interest

The three main elements of the said definition are the primary and secondary interest and their conflict. Primary interests include promoting and protecting the integrity of research method and results and the welfare of patients and research participants. Secondary interests may include not only financial gain but also the desire for professional advancement, recognition for personal achievement, and favors to friends and family or to students and colleagues. The secondary interests are objectionable only when they have greater weight than the primary interest in professional decision-making. For the researcher, financial interests should be subordinate to presenting objective unbiased scientific evidence. The third element, the actual conflict, exists whether or not a particular party is actually influenced by the secondary interest. Both experience and research indicate that under certain conditions

there is a risk that professional judgment may be influenced more by secondary interests than by primary interests [10].

Conflict of interest is a worthwhile subject considering the simple truth that financial ties between industry and academic researchers contain the inherent risk that a guideline development process or research results may be compromised. Conflict of interest policies typically and reasonably focus on financial gain and financial relationships. An “intellectual conflict of interest” is also possible and has been defined as well by Guyatt et al. [9]:

Academic activities that create the potential for an attachment to a specific point of view that could unduly affect an individual’s judgment about a specific recommendation.

Potential conflicts of interest can arise because of the professional position held by the author, such as being employed by a private clinic or sitting on advisory boards related to specific treatments. Researchers can feel pressured to obtain funding, publish, advance careers, or obtain tenure. Universities add to this pressure by pushing researchers to succeed [16]. These personal interests also have the potential to bias researchers when they conflict with research. Furthermore, intellectual COI is further characterized as “important,” when it includes authorship of original studies and peer-reviewed grant funding by the government or nonprofit organizations that directly relate to a recommendation and should be frowned upon, and “less important,” as is the participation in previous guideline panels that must be acknowledged but do not necessarily preclude participation in developing new recommendations. The reason of focus on financial COI is not that financial gains are necessarily more corrupt-

ing than the other interests but that they are relatively more objective and quantifiable [10].

5.3 Recognition and Consequences of Conflict of Interest

Financial conflicts of interest and intellectual bias can influence recommendations for clinical practice. The issue of orthopedic surgeons having industry-related conflicts of interest that could potentially affect the scientific reports that are used to guide orthopedic patient care remains important patient- and community-wise. To begin with, dissemination of orthopedic research can be affected by industry funding-related conflict of interest. Rates of citation may be influenced by characteristics other than the impact factor of the publishing journal. It has been found that high citation rates in medical literature are associated with industry funding and the reporting of an industry-favoring result, among others and in addition to journal impact factor [11]. Even more recently, in a study focused exclusively on orthopedic journals, high level of evidence, large sample size, representation from multiple institutions, and conflict of interest disclosure itself involving a nonprofit organization or for-profit company are associated with higher rates of citation in orthopedics [18]. A possible explanation offered is the possibility that researchers who secure external funding may be able to publish articles that are scientifically superior and therefore more likely to be cited.

Biases from financial and/or intellectual COIs may result from the ability of such conflicts to affect decision-making in a way that is completely hidden from the person making the decision. A related concern involves access to data. In industry-supported research, there is often the case that the investigator may lack full access to the study data. This exposes and creates a new idea of unintentional bias on the researcher’s behalf. Industry sponsors may even go as far as preventing publication of findings not in their favor. Conflict of interest relates to people, not to companies or organizations. Most people

Fact Box 5.2

The three main elements of the definition of a conflict of interest are the primary and secondary interest and their conflict. Secondary interests are objectionable only when they have greater weight than the primary interest in professional decision-making, interest, and their conflict.

involved in research do not recognize the effect of these conflicts on their judgments [5, 23]. Declaring or acknowledging conflicts does not mitigate their effects. Moral licensing occurs when disclosure of a COI reduces feelings of guilt of the advisor, resulting in more biased advice because advisees have been warned [4, 15]. Even though a researcher could be unbiased and declaring no conflict of interest whatsoever, one should not fail to mention the potential conflict of interest of editors and reviewers in relation to the review or publication of research.

Fact Box 5.3

Biases from financial and/or intellectual COIs may result from the ability of such conflicts to affect decision-making in a way that is completely hidden from the person making the decision.

Distinction between an actual conflict of interest, an apparent conflict of interest, and a potential conflict of interest exists. A perceived conflict of interest may be as important as an actual conflict of interest. Financial COI is an important factor impacting the public's trust in research. A perceived conflict of interest can undermine the public's opinion and trust in the quality of public sector services, institutional processes, and public institutions [1].

Fear for biased results or even fake results is legitimate. Some researchers argue that disclosure policies of scientific journals about researchers' financial ties to their work do little to help with bias prevention or identification. Furthermore, valuable research can be discarded by readers if financial ties are declared, even though results were not affected by funding [6]. Another way of affecting readers' perception is bias in control group choice in prospective randomized controlled trials. Such bias can favor products that are made by the company funding the research, and this might be due to the selection of inappropriate comparator products and publication bias. In a recent study looking into possible influences on the findings, protocols,

and quality of drug trials financed by pharmaceuticals, the methodological quality of trials funded by profit organizations was not found to be worse than trials funded through other more independent sources. It was however suggested that protocols were influenced to favor the trial product and results were interpreted more favorably [22]. Another way of looking into this subject is that bias lies in funding decisions made by companies toward studies that are likely to promote their interests and in manipulating the process of research [19]. Finally, as far as perceptions of financial ties to profit organizations by patients, professionals, and researchers, the quality of research evidence is thought to be compromised, and ties should be disclosed in order to enable readers to interpret results through their own scope [14, 17].

5.4 Management of Conflict of Interest

Institutions, professional organizations, and governments establish policies to address the problem of conflict of interest on behalf of the public in order to achieve and maintain high standards of care. Such policies work best when they aim at prevention or mitigation of its impact on research integrity rather than penalty after occurrence [20]. Although penalizing researchers who violate disclosure policies may also assist with prevention, the presence of a conflict of interest does not imply that any individual is improperly motivated. To avoid these and similar mistakes and to provide guidance for formulating and applying such policies, a framework for analyzing conflicts of interest is desirable. Conflict of interest policies should not only address concerns that financial relationships with industry may lead to bias or a loss of trust but should also consider the potential benefits of such relationships in specific situations.

There are three key steps in managing conflict of interest: (a) identifying relevant suspicious relationships, (b) resolving such existing relationships, and (c) disclosing relevant relationships

to learners [7]. Eliminating the conflict may not be possible for ethical or practical reasons.

Requirements for the registration of clinical trials are, in part, a response to concerns about conflict of interest in industry-sponsored research and research reporting. The registration of clinical trials ensures that basic methods for the conduct and analysis of the findings of a study as well as the primary clinical end points to be assessed and reported are specified before the trial begins and before data are analyzed [10]. In the USA, management of COI and objectivity in research is regulated federally by the Department of Health and Human Services [8].

Essentially all medical journals require disclosure of potential conflicts of interest from their authors. Unfortunately the format of that disclosure is not consistent among journals. The management of conflict of interest in academic publishing involves multiple parties: those who research, those who edit and publish, and those who read the research. Awareness of the possibility of conflict of interest is the first step to recognition, and the difficult part in managing this complex issue is declaration of conflict of interest. The responsibility for managing conflict of interest in publishing is equally spread across authors, reviewers, editors, and readers.

For researchers and authors, there is a need to be aware of those situations that have the potential to bias results or perspectives. Potential conflicts may include financial or in-kind support from vested interests. Most clearly and easily identified is research funding sources directly related to the publication, but this support could also include conference scholarships, paid speaking engagements, and prizes. Managing this potential conflict of interest in the first instance involves awareness and acknowledgement of potential or real conflict of interest and declaration on submission. Adherence to sound academic principles, such as being true to the data, not over-claiming, and not under-disclosing, is a protection that can help manage potential conflicts. In several countries, in addition to reporting funding, researchers must disclose COIs in presentations and publications, as well as to colleagues and, in some cases, research subjects [12].

In reviewing manuscripts, reviewers need to hold in mind the potential for conflict of interest and be vigilant for potential conflicts of interest. Reviewers need to check that references used to support statements are being appropriately utilized, that the claims made are reasonable given the findings of the study, that negative findings are fairly reported, and that conclusions and recommendations are commensurate with the implications that can be drawn from the study. Editors also need to consider their own individual conflicts in a similar way to authors and reviewers. They have the job of developing policy and ensuring that authors' and reviewers' interests do not distort publication decisions. It is the readers, the healthcare professionals, that have the most difficult part in this sort of conflict, since they have to develop their own approach to the interpretation of the results. Awareness is the key, and the responsibility is theirs to question whether published material may be influenced by interests of authors, reviewers, or editors [17].

Fact Box 5.4

There are three key steps in managing conflict of interest: (a) identifying relevant suspicious relationships, (b) resolving such existing relationships, and (c) disclosing relevant relationships to learners. Eliminating the conflict may not be possible for ethical or practical reasons.

Take-Home Message

- It is important to understand that having a conflict does not mean wrongdoing.
- A conflict is a situation that increases the potential or risk for bias to occur, and steps should be taken to assure the integrity of the research.
- The goal is to increase transparency, primarily through disclosure and regulation, which allows consumers of the research to consider potential influences related to the conflict.
- Sometimes careful judgment is required based on the facts and the nature of the research.

References

1. Australian Research Council. ARC conflict of interests and confidentiality policy Version 2017.1. <http://www.arc.gov.au/arc-conflict-interest-and-confidentiality-policy-version-2017.1>.
2. Battelle. 2014 Global R&D funding forecast. R&D Magazine. 2013. <https://www.battelle.org/docs/default-source/misc/2014-rd-funding-forecast.pdf?sfvrsn=2>.
3. Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA*. 2003;289:454–65. <https://doi.org/10.1001/jama.289.4.454>.
4. Cain DM, Detsky AS. Everyone's a little bit biased (even physicians). *JAMA*. 2008;299:2893–5. <https://doi.org/10.1001/jama.299.24.2893>.
5. Cosgrove L, Bursztajn HJ, Erlich DR, Wheeler EE, Shaughnessy AF. Conflicts of interest and the quality of recommendations in clinical guidelines. *J Eval Clin Pract*. 2013;19:674–81. <https://doi.org/10.1111/jep.12016>.
6. de Melo-Martín I, Intemann K. How do disclosure policies fail? Let us count the ways. *FASEB J*. 2009;23:1638–42. <https://doi.org/10.1096/fj.08-125963>.
7. Dickerson P, Chappell K. Content integrity, conflict of interest, and commercial support: defining and operationalizing the terms. *J Nurses Prof Dev*. 2015;31:225–30. <https://doi.org/10.1097/NND.0000000000000182>.
8. Electronic Code of Federal Regulations. PART 50—Policies of general applicability. Subpart F—Promoting objectivity in research. <https://www.ecfr.gov/cgi-bin/text-idx?SID=b17a2b836d6343b5c359269f32ed5ccb&mc=true&node=pt42.1.50&rgn=div5#sp42.1.50.f>.
9. Guyatt G, Akl EA, Hirsh J, Kearon C, Crowther M, Gutterman D, et al. The vexing problem of guidelines and conflict of interest: a potential solution. *Ann Intern Med*. 2010;152:738–41. <https://doi.org/10.7326/0003-4819-152-11-201006010-00254>.
10. IOM (Institute of Medicine). Conflict of interest in medical research, education, and practice. Washington, DC: The National Academies Press; 2009.
11. Kulkarni AV, Busse JW, Shams I. Characteristics associated with citation rate of the medical literature. *PLoS One*. 2007;2:e403.
12. Lach HW. Financial conflicts of interest in research: recognition and management. *Nurs Res*. 2014;63:228–32. <https://doi.org/10.1097/NNR.0000000000000016>.
13. Lee K, Bacchetti P, Sim I. Publication of clinical trials supporting successful new drug applications: a literature analysis. *PLoS Med*. 2008;5:e191. <https://doi.org/10.1371/journal.pmed.0050191>.
14. Licurse A, Barber E, Joffe S, Gross C. The impact of disclosing financial ties in research and clinical care: a systematic review. *Arch Intern Med*. 2010;170:675–82. <https://doi.org/10.1001/archinternmed.2010.39>.
15. Loewenstein G, Sah S, Cain DM. The unintended consequences of conflict of interest disclosure. *JAMA*. 2012;307:669–70. <https://doi.org/10.1001/jama.2012.154>.
16. Marcovitch H, Barbour V, Borrell C, Bosch F, Fernández E, Macdonald H, Marusić A, Nylenna M, Esteve Foundation Discussion Group. Conflict of interest in science communication: more than a financial issue. Report from Esteve Foundation Discussion Group, April 2009. *Croat Med J*. 2010;51:7–15.
17. O'Brien L, Lakeman R, O'Brien A. Managing potential conflict of interest in journal article publication. *Int J Ment Health Nurs*. 2013;22:368–73. <https://doi.org/10.1111/j.1447-0349.2012.00869.x>.
18. Okike K, Kocher MS, Torpey JL, Nwachukwu BU, Mehlman CT, Bhandari M. Level of evidence and conflict of interest disclosure associated with higher citation rates in orthopedics. *J Clin Epidemiol*. 2011;64:331–8. <https://doi.org/10.1016/j.jclinepi.2010.03.019>.
19. Resnik DB, Elliott KC. Taking financial relationships into account when assessing research. *Account Res*. 2013;20:184–205. <https://doi.org/10.1080/08989621.2013.788383>.
20. Resnik DB. Science and money: problems and solutions. *J Microbiol Biol Educ*. 2014;15:159–61. <https://doi.org/10.1128/jmbe.v15i2.792>.
21. Rockey SJ, Collins FS. Managing financial conflict of interest in biomedical research. *JAMA*. 2010;303:2400–2. <https://doi.org/10.1001/jama.2010.774>.
22. Schott G, Pacht H, Limbach U, Gundert-Remy U, Ludwig WD, Lieb K. The financing of drug trials by pharmaceutical companies and its consequences. Part 1: a qualitative, systematic review of the literature on possible influences on the findings, protocols, and quality of drug trials. *Dtsch Arztebl Int*. 2010;107:279–85. <https://doi.org/10.3238/arztebl.2010.0279>.
23. Thagard P. The moral psychology of conflicts of interest: insights from affective neuroscience. *J Appl Philos*. 2007;24:367–80. <https://doi.org/10.1111/j.1468-5930.2007.00382.x>.



Ethics in Clinical Research

6

Naomi Roselaar, Niv Marom, and Robert G. Marx

6.1 Introduction

6.1.1 Tuskegee Syphilis Experiment

In 1932, the US Public Health Service and Tuskegee Institute in Alabama began an observational study of syphilis in African American men [6]. Called the “Tuskegee Study of Untreated Syphilis in the Negro Male,” the study was intended to demonstrate the need for a syphilis treatment program [6]. Approximately 600 subjects, of whom 400 had syphilis, were told nothing of their disease. Despite the availability of bismuth, arsenic, mercury, and later penicillin, as therapy, the subjects were offered no treatment [11]. Subjects suspected of receiving injections of arsenic or mercury were immediately replaced [32]. As reported in a paper read before the 14th Annual Symposium on Recent Advances in the Study of Venereal Diseases in January 1964, “Fourteen young, untreated syphilitics were added to the study to compensate for this” [32]. Following media outrage, the study concluded in 1972 when a nine-person panel found that no information had been provided to subjects before they agreed to participate [6]. This 40-year experiment on non-consenting, medically neglected

subjects became the longest nontherapeutic study of humans in medical history [11]. The Tuskegee Syphilis Experiment illustrated the exploitation of vulnerable patients, the need for informed consent, and the misrepresentation of minority populations in clinical studies. Since then, great care has been taken at all levels of clinical research to employ ethical guidelines and regulatory committees to oversee studies involving human subjects.

6.2 Regulatory and Ethical Guidelines in Clinical Research

When proposing and conducting experiments involving human subjects, researchers must comply with international, federal, and institutional guidelines to protect participants. The US Department of Health and Human Services’ “Common Rule,” the Institutional Review Boards (IRBs) at individual institutions, and the Health Insurance Portability and Accountability Act of 1996 (HIPAA) are three primary regulatory measures for ethics in clinical research. All three were enacted after the 1964 adoption of the World Medical Association’s *Declaration of Helsinki: Ethical Principles for Medical Research Involving Human Subjects*.

N. Roselaar · N. Marom · R. G. Marx (✉)
Orthopedic Surgery, Hospital for Special Surgery,
New York, NY, USA
e-mail: roselaarn@hss.edu; maromn@hss.edu;
marxr@hss.edu

6.2.1 Declaration of Helsinki

The Declaration of Helsinki (DOH) has been considered the gold standard for ethics in clinical research [26]. In its current form, the DOH applies to human subjects, data, and material (WMA). Central tenets of the DOH protect the health and rights of all patients involved in clinical research and advocate for the continuous evaluation of safety, effectiveness, efficiency, accessibility, and quality of human subject research [46]. It was originally composed of 14 short statements outlining ethical guidelines for conducting human subject research [47]. Since its inception, the DOH has been revised seven times, 1975 (Tokyo), 1983 (Venice), 1989 (Hong Kong), 1996 (Somerset, South Africa), 2000 (Edinburgh), 2008 (Seoul), and 2013 (Fortaleza, Brazil), and revised twice [46]. By 2014, the DOH included 37 detailed principles [46]. The basis of the declaration stems from the Nuremberg Code [5]. This seminal code of ethics was established at the conclusion of Nuremberg trials for Nazi war crimes, including horrifically violent human medical experiments on Holocaust victims [34].

6.2.2 Common Rule

In the United States, the Department of Health and Human Services (HHS) issues regulations on the ethical conduct of research on humans [22]. The HHS *Code of Federal Regulations 45 Part 46 Protection of Human Subjects* was developed in 1981 and updated in 2009 [40]. The policy is known colloquially as the “Common Rule” and protects human subjects in research conducted or supported by a federal department or agency [40]. It requires researchers to provide informed, written consent, full disclosure of the benefits and foreseeable risks of the proposed study, and a statement addressing subject rights to refuse participation at any point [40]. Considered vulnerable populations, pregnant women, human fetuses (definite as the “product of conception from implantation until delivery”), neonates, prisoners, and children are offered additional protec-

tions under the Common Rule [40]. Revisions to the Common Rule were proposed in 2015 by HHS and 15 other federal departments and agencies [22]. The updates are designed to reflect changes in research over the 35 years since the inception of the Common Rule [22]. Goals include enhancing respect and strengthening informed consent, particularly for the long-term use of de-identified biospecimens; enhancing safeguards by specifying privacy and security measures concerning identifiable information; streamlining Institutional Review Board (IRB) review by clarifying levels of risk and the IRB process for multisite studies; and calibrating oversight [40]. The HHS offered the opportunity for public comments on revisions until January 2016 [42].

6.2.3 Institutional Review Board

The Common Rule also states that research conducted or supported by organizations outside a federal department or agency must be compliant with the host’s Institutional Review Board (IRB) [40]. Like the Common Rule, IRBs aim to provide ethical and regulatory oversight for research with human subjects. On an institutional level, they ensure compliance with external laws, policies, and regulations [27]. Both the Common Rule and IRBs operate to uphold ethical principles defined in the Belmont Report of 1979 [40, 27]. The primary principles include respect for persons, beneficence, and justice [27]. Research approved by an IRB is subject to annual continuing reviews.

6.2.4 HIPAA

In addition to the protection of subjects themselves, information that identifies subjects must also be protected. The Health Insurance Portability and Accountability Act of 1996 seeks to increase privacy protection for human research participants [10]. This includes the privacy and security of information that could be used to identify subjects in a particular study such as name, medical record number, birthdate, social

security number, address, or identifying photograph. As electronic medical records become more prevalent, and security and privacy issues extend to the online storage of identifiable data, regulations must also change. In 2000, 2004, 2009, and 2013, HIPAA has been modified and extended to reflect technological advances [41].

The HIPAA Privacy Rule, which protects identifiable health information, was introduced in 2000 and mandated nationally in 2003 [29]. By protecting ownership and transfer of specific protected health information (PHI), the Privacy Rule aims to safeguard health information associated with individuals while facilitating data flow to maximize the quality of health care [43]. PHI comprises any information that can be used to identify a study participant [43]. The implementation of the HIPAA Privacy Rule in clinical research has also provoked criticism. In 2007, JAMA published a study in which more than two-thirds of surveyed epidemiologists perceived “substantial, negative influence on the conduct of human subjects health research” after the implementation of the HIPAA Privacy Rule [29]. To encompass the protection of electronic PHI (e-PHI), the HIPAA Security Rule (HSR) was enacted in 2003 [44].

6.3 Ethical Population Representation

The lack of diverse racial and ethnic representation in clinical research prevents the best possible treatment for disease outcomes in heterogeneous populations [30]. To address this issue, Congress passed the National Institute of Health (NIH) Revitalization Act of 1993, intended to catalyze the diversification of participants in clinical research [28]. With minimal exceptions the Revitalization Act requires NIH-funded clinical research to include women and members of minority groups [28].

Twenty years after the introduction of regulatory laws to diversify participation in clinical research, the minority participation in cancer clinical trials remained minimal [7]. The lack of minority representation persists across many

other fields in clinical research including cardiovascular disease, respiratory disease, mental health services, and substance abuse [2, 24, 33, 3].

Despite legislation, barriers in inclusion criteria may prevent trials from including minority populations. English language fluency is required for clinical trials more and more frequently [16]. The mistrust of healthcare professionals and the lack of understanding of clinical research also contribute to low rates of minority participation [2]. Decreased access to academic institutions pursuing clinical research by minorities influences recruitment diversity [14].

6.4 Ethical Publishing Practices

6.4.1 Data Fraud and Misconduct

One of the most highly publicized fraudulent studies is Dr. Andrew Wakefield’s case series suggesting an association between the MMR vaccine and autism [45]. Published in *The Lancet*, a British medical journal, in 1998, Dr. Wakefield’s paper caught the attention of mainstream media [31]. Consequently, the rate of MMR vaccines for toddlers in the United Kingdom decreased from 83.1% in 1997 to 69.9% in 1998 [39]. A one-page commentary titled, “Retraction of an Interpretation” was published in 2004 by 10 of the 13 authors of the original article [25]. Simultaneously, editors at *The Lancet* acknowledged a lack of financial disclosures by Dr. Wakefield et al. and reaffirmed “the paper’s suitability, credibility, and validity for publication” [21]. In 2010, *The Lancet* retracted Dr. Wakefield’s article.

This example demonstrates many important facets of the ethics of clinical research. The researchers failed to report accurate findings and drew speculative conclusions from a small, non-representative case series [31]. These unethical actions by the authors were compounded by irresponsible publishing practices at *The Lancet*. The publishers failed to require proper disclosure of conflicts of interest, specifically those that revealed Dr. Wakefield’s financial gains related to the research conclusions [12]. Furthermore, *The Lancet* only retracted the fraudulent article

12 years after its initial publication [15]. After the retraction of the article, investigative journalist Brian Deer published multiple articles in the BMJ revealing Dr. Wakefield's connection to lawyer Richard Barr [12]. Barr, who was working to file a lawsuit against vaccine manufacturing companies, provided Dr. Wakefield with £400,000 through the Legal Aid Fund while also representing the anti-vaccine organization Justice, Awareness, and Basic Support (JABS) [13]. Barr used his connection to JABS to find patients for Dr. Wakefield's study [12].

However, cases such as Dr. Wakefield's are not common. Although data fraud is difficult to monitor and likely underreported, confirmed cases of data fabrication, falsification, and plagiarism exist among 0.01% of scientists according to the US Public Health Service [19]. Dr. Wakefield's willful deceit through data falsification is classified as fraud, while misconduct refers to honest errors in ethical research practices [19]. In addition to data fraud and misconduct, there are many other aspects of clinical research for which good clinical practices must be observed. They include conflict of interest disclosure, self-citation, and distinguishing predatory journals.

6.4.2 Conflict of Interest

Conflicts of interest in clinical orthopedic research are any instances of overlapping personal, financial, or academic involvement that may bias or influence a participant's work. Investigators are required to submit conflict of interest statements for project proposals, manuscript publications, and presentations at conferences. This benefits consumers of the research as it provides context for the circumstances under which research is conducted. It is the responsibility of authors, editors, peer reviewers, and another other staff members who play a role in determining the publication or presentation of a study to disclose any relevant conflicts of interest [23]. Internationally, many orthopedics journals follow the "Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals" established by the International Committee of Medical

Journal Editors (ICMJE). The conflicts of interest included by the ICJME conflict of interest form include financial activities related to the work as well as relevant financial activities outside the submitted work. Relevant financial activities may include relationships with a pertinent entity such as a government agency, foundation, academic institution, or commercial sponsor; grants; personal fees; royalties; leadership positions; and nonfinancial support [23].

6.4.3 Self-Citation

Self-citation refers to referencing an article from the same journal [8]. The rate of self-citation for a medical journal is defined by the number of self-citations divided by the number of total references made by that journal in a specified time period [20]. In orthopedics journals, many factors influence the differences in self-citation rates. Self-citation rates are highest in sub-specialized journals due to their specificity [36]. Specialized orthopedics journals including *Spine*, *Arthroscopy*, and *FAI* have self-citation rates two and three times higher than the general orthopedics journals *CORR* and *JBJS (Am)*, respectively [36]. Rates of self-citation are categorized as "high" if they are at or above 20% (JCR).

Self-citation rates are relevant in the calculation of medical journal impact factors [17]. According to the Science Citation Index, the impact factor for medical journals "measures the average number of citations received in a particular year by papers published in the journal during the two preceding years" [8]. Practices surrounding the manipulation of self-citation rates introduce bias into impact factors [20]. For journals in which self-citations dominate the references, the true contribution to the journal's discipline may be misrepresented [18].

6.4.4 Predatory Journals

An increase in the existence of open-access publications with minimal or no peer review has been accredited to the rise of spam emails [4]. Known

as “predatory journals” [1], these publications often require authors to pay high fees for the processing and peer-review process with no follow-through [4]. Jeffrey Beall, a former Scholarly Initiatives Librarian at the University of Colorado Denver coined the term and proposed the first list of predatory journals [1]. He warned that these publishers “exploit the author-pays model, damage scholarly publishing, and promote unethical behavior by scientists” [1]. In 2017 Beall’s list of predatory journals, which had been used as a government standard, was removed from his webpage [38]. However, institutions such as the Yale University Library system continue to recommend it and other similar lists [48].

Articles submitted to predatory journals are published quickly due to the limited or nonexistent review process [35]. Additionally, published articles are often non-indexed despite advertisements to the contrary [9]. Non-indexed articles cannot be retrieved through an online search [35].

Human behavioral scientists in Poland aimed to shed light on the issue of predatory journals through a systematic study in which they created a profile for an imaginary scientist and applied for editorial positions at 360 journals [37]. False online accounts, journal and book publications, and faculty positions—none of which could be verified—were compiled into the fake application. Of the 360 editorial positions to which the fake application was submitted, 120 were for journals indexed by Journal Citation Reports (JCR), 120 were listed on the Directory of Open Access Journals (DOAJ), and 120 were included in Beall’s list of predatory journals. Acceptances came from 40 journals included on Beall’s list and 8 journals listed on the DOAJ [37]. Fittingly, the fake editor’s name was Dr. Anna O. Szust—similar to *Oszust*, the Polish word for “fraud.” All offers for editorial positions were declined [37].

Take-Home Message

- Ethical considerations are important at many levels and processes in clinical research.
- Investigators, ethical review board members, publishers, and peer reviewers all contribute responsibility to maintaining high ethical standards in clinical research.

References

1. Beall J. Predatory publishers are corrupting open access: journals that exploit the author-pays model damage scholarly publishing and promote unethical behaviour by scientists, argues Jeffrey Beall. *Nature*. 2012;489(7415):179.
2. Boden-Albala B, Carman H, Southwick L, Parikh NS, Roberts E, Waddy S, et al. Examining barriers and practices to recruitment and retention in stroke clinical trials. *Stroke*. 2015;46(8):2232–7.
3. Burlew K, Larios S, Suarez-Morales L, Holmes B, Venner K, Chavez R. Increasing ethnic minority participation in substance abuse clinical trials: lessons learned in the National Institute on Drug Abuse’s Clinical Trials Network. *Cult Divers Ethn Minor Psychol*. 2011;7(4):345–56.
4. Butler D. The dark side of publishing. *Nature*. 2013;495:433–5.
5. Carlson RV, Boyd KM, Webb DJ. The revision of the declaration of Helsinki: past, present and future. *Br J Clin Pharmacol*. 2004;57(6):695–713.
6. Centers for Disease Control and Prevention. The Tuskegee timeline [Internet]. U.S. Public Health Service Syphilis Study at Tuskegee. 2017 [cited 2017 Oct 21]. <https://www.cdc.gov/tuskegee/timeline.htm>.
7. Chen MS, Lara PN, Dang JHT, Paterniti DA, Kelly K. Twenty years post-NIH Revitalization Act: enhancing minority participation in clinical trials (EMPaCT): laying the groundwork for improving minority clinical trial accrual: renewing the case for enhancing minority participation in cancer clinical trials. *Cancer*. 2014;120(Suppl 7):1091–6.
8. Clarivate Analytics. Journal Impact Factor [Internet]. InCites Help. [cited 2017 Dec 19]. <http://ipscience-help.thomsonreuters.com/inCites2Live/indicatorsGroup/aboutHandbook/usingCitationIndicatorsWisely/jif.html>.
9. Clark J. How to avoid predatory journals—a five point plan. *BMJ Opin*. 2015. <https://blogs.bmj.com/bmj/2015/01/19/jocelyn-clark-how-to-avoid-predatory-journals-a-five-point-plan/>.
10. 104th Congress. Health Insurance Portability and Accountability Act of 1996 [Internet]. U.S. Department of Health and Human Services. 1996 [cited 2017 Oct 21]. <https://aspe.hhs.gov/report/health-insurance-portability-and-accountability-act-1996>.
11. Corbie-Smith G. The continuing legacy of the Tuskegee Syphilis Study: considerations for clinical investigation. *Am J Med Sci*. 1999;317(1):5–8.
12. Deer B. How the vaccine crisis was meant to make money. *BMJ*. 2011;342:c5258.
13. Deer B. Revealed: secret payments to MMR doctor Wakefield at heart of vaccine crusade. 2006. brian-deer.com.
14. Durant RW, Wenzel JA, Scarinci IC, Paterniti DA, Fouad MN, Hurd TC, et al. Perspectives on barriers and facilitators to minority recruitment for clinical trials.

- cal trials among cancer center leaders, investigators, research staff, and referring clinicians: enhancing minority participation in clinical trials (EMPaCT). *Cancer*. 2014;120(Suppl 7):1097–105.
15. Eggertson L. Lancet retracts 12-year-old article linking autism to MMR vaccines. *CMAJ*. 2010;182(4):e199–200.
 16. Egleston BL, Pedraza O, Wong YN, Dunbrack RL, Griffin CL, Ross EA, et al. Characteristics of clinical trials that require participants to be fluent in English. *Clin Trials*. 2015;12(6):618–26.
 17. Frandsen TF. Journal self-citations -analysing the JIF mechanism. *J Informetr*. 2007;1(1):47–58.
 18. Garfield E. Journal self-citation in the Journal Citation Reports – Science Edition [Intranet]. Clarivate Analytics. 2002 [cited 2018 Nov 13]. <https://clarivate.com/essays/journal-self-citation-jcr/>.
 19. George SL, Buyse M. Data fraud in clinical trials. *Clin Invest (Lond)*. 2015;5(2):161–73.
 20. Hakkalamani S, Rawal A, Hennessy MS, Parkinson RW. The impact factor of seven orthopaedic journals. *J Bone Joint Surg [Br]*. 2006;88(2):159–62.
 21. Horton R, Murch S, Walker-Smith J, Wakefield A, Hodgson H. A statement by the editors of the lancet. *Lancet*. 2004;363:P820–1.
 22. Hudson KL, Collins FS. Bringing the common rule into the 21st century. *N Engl J Med*. 2015;373(24):2293–6.
 23. ICMJE. Recommendations for the conduct, reporting, editing, and publication of scholarly work in medical journals. 2017.
 24. McGarry ME, McColley SA. Minorities are under-represented in clinical trials of pharmaceutical agents for cystic fibrosis. *Ann Am Thorac Soc*. 2016;13(10):1721–5.
 25. Murch S, Anthony A, Casson D, Malik M, Mark B, Dhillon A, et al. Retraction of an interpretation. *Lancet*. 2004;363:750.
 26. Muthuswamy V. The new 2013 seventh version of the declaration of Helsinki—more old wine in a new bottle? *Indian J Med Ethics*. 2014;11(1):2–4.
 27. National Institute of Environmental Health Sciences. Institutional Review Board [Internet]. National Institutes of Health. 2015 [cited 2017 Oct 21]. <https://www.niehs.nih.gov/about/boards/irb/index.cfm>.
 28. National Institutes of Health. NIH policy and guidelines on the inclusion of women and minorities as subjects in clinical research [Internet]. Office of Extramural Research. 2017 [cited 2017 Dec 12]. https://grants.nih.gov/grants/funding/women_min/guidelines.htm.
 29. Ness RB. Influence of the HIPAA privacy rule on Health Research. *JAMA*. 2007;298(18):2164–70.
 30. Oh SS, Galanter J, Thakur N, Pino-Yanes M, Barcelo NE, White MJ, et al. Diversity in clinical and biomedical research: a promise yet to be fulfilled. *PLoS Med*. 2015;12(12):e1001918.
 31. Rao T, Andrade C. The MMR vaccine and autism: sensation, refutation, retraction, and fraud. *Indian J Psychiatry*. 2011;53(2):95–6.
 32. Rockwell DH, Yobs AR, Moore B Jr. The Tuskegee study of untreated syphilis: the 30th year of observation. *Arch Intern Med*. 1964;114(6):792–298.
 33. Santiago CD, Miranda J. Progress in improving mental health services for racial-ethnic minority groups: a ten-year perspective. *Psychiatr Serv*. 2014;65(2):180–5.
 34. Shuster E. Fifty years later: the significance of the Nuremberg code. *N Engl J Med*. 1997;337:1436–40.
 35. Shyam A. Predatory journals: what are they? *J Orthop Case Rep*. 2015;5(4):1–2.
 36. Siebelt M, Siebelt T, Pilot P, Bloem RM, Bhandari M, Poolman RW. Citation analysis of orthopaedic literature; 18 major orthopaedic journals compared for impact factor and SCImago. *BMC Musculoskelet Disord*. 2010;11(4):1–7.
 37. Sorokowski P, Kulczycki E, Sorokowska A, Pisanski K. Predatory journals recruit fake editor. *Nature*. 2017;543:481–3.
 38. Strielkowski W. Predatory journals: Beall's list is missed. *Nature*. 2017;544:416.
 39. Thomas DR, Salmon RL, King J. Rates of first measles-mumps-rubella immunisation in Wales (UK). *Lancet*. 1998;351:1927.
 40. U.S. Department of Health and Human Services. Code of Federal Regulations Title 45 Part 46 [Internet]. Office for Human Research Protections. 2009 [cited 2017 Oct 21]. <https://www.hhs.gov/ohrp/regulations-and-policy/regulations/45-cfr-46/index.html#subparta>.
 41. U.S. Department of Health and Human Services. HIPAA for Professionals [Internet]. Office for Civil Rights. 2017 [cited 2017 Oct 21]. <https://www.hhs.gov/hipaa/for-professionals/index.html>.
 42. U.S. Department of Health and Human Services. NPRM for revisions to the common rule. Federal Register. 2015.
 43. U.S. Department of Health and Human Services. Summary of the HIPAA Privacy Rule [Internet]. Office for Civil Rights. 2013 [cited 2017 Oct 21]. <https://www.hhs.gov/hipaa/for-professionals/privacy/laws-regulations/index.html>.
 44. U.S. Department of Health and Human Services. Summary of the HIPAA Security Rule [Internet]. Office for Civil Rights. 2013 [cited 2017 Oct 21]. <https://www.hhs.gov/hipaa/for-professionals/security/laws-regulations/index.html>.
 45. Wakefield A, Murch S, Anthony A, Linnell J, Casson D, Malik M, et al. RETRACTED: ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet*. 1998;351:P637–41.
 46. World Medical Association. WMA Declaration of Helsinki—ethical principles for medical research involving human subjects. Helsinki; 1964.
 47. World Medical Organization. Declaration of Helsinki (1964). *Br Med J*. 1996;313(7070):1448–9.
 48. Yale University Library. Choosing a journal for publication of an article: list of suspicious journals and publishers [Internet]. 2018 [cited 2017 Dec 19]. <https://guides.library.yale.edu/c.php?g=296124&p=1973764>.

Part II

How to Get Started with Clinical Research?



How to Get Started: From Idea to Research Question

7

Lachlan M. Batty, Timothy Lording,
and Eugene T. Ek

7.1 Research Questions and the Decisions We Make Every Day

In trying to achieve the best possible outcomes for patients, orthopedic surgeons are faced with multiple decisions every day. Evidence-based medicine (EBM), as described by Sackett, is “the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients” [17]. It is a systematic approach to optimize patient care and dictates that our practice as surgeons is guided by data to support each of these decisions.

Orthopedic surgeons are an inherently inquisitive group, making observations, continually refining practice, and adopting novel and innovative technologies and techniques. In this context, many decisions made by surgeons are based on a limited evidence base or by extrapolation of

orthopedic principles. It is important for these decisions to be continually challenged and objectively evaluated in the clinical environment to account for known and unknown variables. Testing these advances and reforms as structured and answerable research questions is a fundamental initial step in the process of conducting clinical research. With this approach the surgeon-scientist uses scientific principles to guide practice and practice to guide science with the ultimate aim of optimizing patient outcomes.

7.2 The Importance of a Research Question

Developing a clinically relevant, structured, and answerable research question that addresses a knowledge gap is a prerequisite in conducting clinical research. The importance of the research question cannot be understated as it can impact the design, length, cost, and feasibility of the study. The research question is critical in determining inclusion and exclusion criteria, ultimately impacting the extrinsic validity of the study. For example, if a study investigates a group of patients at high risk for a certain condition, subsequent extrapolation to a wider population may not be appropriate. The dependent and independent variables are also shaped early on by the research question. Hulley et al. [10] describe the “FINER criteria” to aid in the development of a research question. By these criteria, a proposed study should be *Feasible*,

L. M. Batty (✉)
Department of Orthopaedics, St Vincent’s Hospital
Melbourne, Melbourne, VIC, Australia

Department of Orthopaedics, Western Health,
Melbourne, VIC, Australia

T. Lording
Melbourne Orthopaedic Group,
Melbourne, VIC, Australia

E. T. Ek
Melbourne Orthopaedic Group,
Melbourne, VIC, Australia

Department of Surgery, Monash Medical Centre,
Monash University, Melbourne, VIC, Australia

Interesting, Novel, Ethical, and Relevant. Each of these elements should be met when developing an idea into a research question as failure to recognize any one element may impede progression of any investigation.

7.3 Types of Research Questions

Research questions can take many forms; however, many authors divide them into four main categories [9, 11]. These pertain to (1) the safety and efficacy of an intervention, (2) the etiology of a pathological process, (3) the diagnosis of a condition or pathological process, or (4) the prognosis of a condition.

Investigating an “intervention” may include an operation, modification to an operation, a medication, or other forms of treatment, such as a rehabilitation program. Examples of research questions assessing interventions include:

- Does combined lateral extra-articular tenodesis reduce graft rupture rates compared with isolated intraarticular ACL reconstruction in high-risk patients? [7]
- Do selective COX-2 inhibitors affect pain control and healing after arthroscopic rotator cuff repair? [15]
- Does operative management of acute Achilles tendon rupture result in a reduced re-rupture rate compared to non-operative management with an accelerated rehabilitation program? [21]

Investigating the etiology of a pathological process aims to determine the causes or risk factors for a pathological process. Potential causes are multiple, including traumatic, degenerative, and genetic. Unlike investigating an intervention, factors investigated in etiological studies are not typically controlled by the surgeon, and therefore study designs generally differ. Examples of research questions that investigate the etiology of a condition include:

- Is ACL rupture a risk factor for development of osteoarthritis of the knee in male soccer players? [20]

- What are the risk factors for infection after shoulder arthroscopy? [3]

Studies investigating the utility of diagnostic modalities of a condition or pathological process may include evaluation of clinical examination techniques and biochemical, hematological, pathological, or radiological investigations. Some examples include:

- Following medial opening wedge high tibial osteotomy, does high-resolution computer tomography have higher sensitivity in diagnosis of lateral hinge fracture compared to plain radiography? [13]
- What are the sensitivity and specificity of the Lachman, anterior drawer, and pivot shift tests for clinical diagnosis of anterior cruciate ligament rupture? [2]
- Can preoperative magnetic resonance imaging predict irreparable rotator cuff tears? [12]

Prognostication is important for allowing communication of expectations to the patient regarding outcomes, whether it be following non-operative or operative intervention. In some cases, these are similar to etiological investigations if the outcome of interest is a secondary pathology. For example:

- What is the midterm risk for rotator cuff tear arthropathy progression in patients with an asymptomatic rotator cuff tear? [4]
- What are the return-to-sport rates after anterior cruciate ligament reconstructive surgery? [1]
- Is anterior cruciate ligament reconstruction effective in preventing secondary meniscal tears and osteoarthritis? [18]

7.4 Identifying Knowledge Gaps

When there are insufficient or inadequate data to adequately guide the clinical decision-making process, there is a “knowledge gap.” When developing an idea into a research question, it must firstly be confirmed that the question has not pre-

viously been answered satisfactorily or entirely. This confirms a knowledge gap. Before embarking on a lengthy and time-consuming investigation, it is important for researchers to evaluate the existing literature. The depth of a literature review can vary, ranging from an ad hoc database search and narrative review to undertaking a structured systematic review or meta-analysis of a particular topic, which is in itself a type of publishable study addressed in a later chapter in this book.

Inherently, a systematic approach to reviewing the literature will lead to a more complete and thorough evaluation of the available data and therefore guide the prospective researcher's questions to greater depth and completion. Often, this goes beyond what is necessary in preparing a research question however, and indeed, to undertake a systematic review, one must start out with a research question to begin with in order to hone relevant articles on the subject.

Often, a research question will gestate from a clinical experience or dilemma, and the aspiring researcher will approach the relevant literature to see what information already exists. If it does not already exist, then a knowledge gap can be established. Even if studies do exist for a particular research question, there may still be substantial knowledge gaps or conflict among reported results. The researcher should consider whether the studies could be improved upon, for example, with greater patient numbers or more controlled methodology and reporting of outcomes.

One example, of many, that highlights the role of systematic review in this setting, is the ongoing STABILITY study [7, 8]. Such rigorous reviews of existing literature may also be beneficial in obtaining funding for a research project and are frequently of publishable quality as systematic reviews. Identification of planned and ongoing investigations is also important within systematic reviews and meta-analyses. With recent requirements for investigators to prospectively register clinical trials, searching trial registry databases [6] is a way of identifying these. PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) provides evidence-based "gold-standard" guidelines for reporting in systematic reviews and meta-analyses.

7.5 The PICO Approach

The PICO (Patient, Intervention, Comparison, Outcome) approach is a well-accepted method of developing a clinical research question [16, 19]. The structured approach guides consideration of important elements of the research question individually. The framing of research questions using this format has been shown to be independently associated with better overall reporting quality in randomized trials [16]. The key is specificity in definitions. Some authors add the element of Time or Timeframe to the acronym making PICOT [16]. This approach lends itself particularly well to investigating the safety and efficacy of an intervention; however, similar principles can be applied when developing other types of research questions.

7.5.1 Patient

The patient population reflects the inclusion and exclusion criteria of a subsequent study. This impacts the extrinsic validity of the study (i.e., who the study results can be applied to) and will affect recruitment rates, event rates, and ethics applications. For example, liberal inclusion criteria may accelerate recruitment rates, whereas focusing on a high-risk population for a condition may increase the event rate. Defining terms such as "high risk" is important and can be elucidated from the initial literature review identifying the knowledge gap. Inclusion of minors may have implications for ethics approval or informed consent process.

Factors to consider in defining the patient group may include demographic factors (e.g., age, sex), skeletal maturity (skeletal immature or mature), nature of the intervention (revision or primary procedure), and factors that may impact on treatment outcome such as past medical history and other conditions (e.g., ligamentous laxity).

The patient population may be further characterized by the pathological process. In an investigation assessing the efficacy of meniscal root repair, for example, patients with traumatic and degenerative

tears may be considered differently. Investigating the efficacy of a new biological agent for chondral restoration may not be applicable to all patients with knee degeneration. Considering the underlying diagnoses of osteoarthritis, inflammatory or post-traumatic arthritis may be warranted as these have different underlying disease processes.

In a randomized control trial, the randomization process aims to evenly distribute known and unknown variables between treatment groups, eliminating confounding; however, defining the patient population sets the boundaries for who is included in this process and therefore who the outcomes can be applied to.

7.5.2 Intervention

The intervention to be investigated should be defined in detail, adhering to the scientific principle of reproducibility. The aim is to allow others to reproduce the investigation or apply the intervention to their own patients. Required details in the descriptions of an intervention will depend if it is a procedure, medication, or treatment program. In terms of a procedure, it should include a detailed description of the pre-, intra-, and postoperative course. Specifics may include things like the position of the limb for graft tensioning and rehabilitation protocols. Detailing who can perform the procedure may also be important; procedures performed by both residents and attending surgeons may add a confounding variable that requires consideration when evaluating the intervention. On the other hand, this may more accurately reflect real-world applicability and is a good example of the importance of defining what the researcher wishes to achieve in asking their research question. In terms of medications, specific details must also be listed. For example, an investigation into platelet-rich plasma (PRP) should clearly detail the preparation of the formulation including if it is leukocyte rich or poor and the number of and interval between injections.

7.5.3 Comparison

The comparison should be described in the same level of detail as the intervention. The comparator may be a placebo in what is known as a “placebo-controlled” study, or the comparator may be an alternative intervention in what is known as a “head-to-head” study. An example of the latter might include comparing differing operative management and operative versus non-operative management, such as physiotherapy, rehabilitation, or bracing.

Sometimes, the comparator can be an alternative patient population in itself, where the intervention is kept constant, but the patient population differs. An example of such a study might be the assessment of rotator cuff repair failure rates between patients under 70 years of age compared to adults over 70 years of age.

7.5.4 Outcome

The outcome of the trial represents the dependent variable in the study. There can be more than one outcome; however, there should be at least one primary outcome. This can take a number of forms including radiological, clinical, or patient-reported outcome measures. Terms such as “failure” need to be clearly defined; does this constitute a revision procedure, graft rupture, ongoing symptoms, or something less tangible, such as patient dissatisfaction? The method for diagnosis or diagnostic criteria for the defined outcome should be detailed (i.e., clinical examination, radiological assessment, or other assessment methods with strict definitions as to what defines a positive result). A time frame should also be included in the research question; for example, investigating anterior cruciate ligament failure rates will yield different results at 6 months compared to 6 years.

7.6 Hypothesis Testing

Following development of a research question, hypotheses are defined that the authors can aim to accept or reject, following conduction of a rigorous investigation.

Hypothesis should be made prior to the start of the investigation. The null hypothesis provides that there is no difference between the groups and the intervention is not superior to the control condition. A hypothesis can be one-sided or two-sided. A one-sided hypothesis states a direction associated with a difference between groups [5]. For example, a one-sided hypothesis may state that patient-reported outcome scores are higher in patients undergoing hamstring ACL reconstruction compared to those having patella tendon graft. A two-sided hypothesis would be that there is a difference in patient-reported outcomes between hamstring and patella tendon ACL reconstruction. In this later case, the patient-reported outcome measures in the hamstring group could be either higher or lower.

Clinical Vignettes/Case Examples

Case Example: The STABILITY Study

Clinical Scenario

Graft re-rupture following ACL reconstruction remains a concern, especially in young patients returning to sports. Augmentation of ACL reconstruction with a lateral extra-articular tenodesis (LET) procedure is one potential strategy to reduce the failure rate.

Identification of the Knowledge Gap

A review of the literature using a systematic approach confirmed a statistically significant reduction in pivot shift in favor of the combined lateral extra-articular tenodesis in conjunction with ACL reconstruction [8]. The available data was inconclusive, however, due to inadequate internal validity, sample size, methodological consistency, and variable standardization of protocols and outcomes. The investigators designed a study to investigate whether or

not combining extra-articular tenodesis reduces the risk of graft failure in a high-risk population [14].

PICO Approach

- Patient
 - Skeletally mature male and female patients aged 14–25 with an ACL deficient knee. High risk for graft rupture was defined as two or more of competitive pivoting sports, grade two pivot shift or greater, and generalized ligament laxity (Beighton score of four or greater). Exclusion criteria include previous ACL reconstruction on either knee, multi-ligament injury (two or more ligaments requiring surgical attention), symptomatic articular cartilage defect requiring treatment other than debridement, greater than 3° of asymmetric varus, or inability to complete outcome questionnaires
- Intervention
 - Lateral extra-articular tenodesis in addition to standard ACL reconstruction as described in the Comparison section. The LET is a modified Lemaire procedure where a 1-cm-wide × 8-cm-long strip of iliotibial band is fashioned, leaving Gerdy's tubercle attachment intact. The graft is tunneled under the fibular collateral ligament (FCL) and attached to the femur with a staple distal to the intermuscular septum and just proximal to the femoral insertion of the FCL. Fixation is performed with the knee at 70° flexion, neutral rotation. Minimal tension is applied to the graft. The free end is then looped back onto itself and sutured using the No. 1 Vicryl
- Comparison
 - Anatomic ACL reconstruction using a four-strand autologous ham-

string graft. If the diameter of the graft is found to be less than 7.5 mm, semitendinosus will be tripled (five-strand graft) providing a greater graft diameter. Femoral tunnels will be drilled using an antero-medial portal technique, with suspensory femoral fixation. Tibial fixation will be provided by interference screw

- Outcome
 - Primary: graft failure at 24 months. This is defined as symptomatic instability requiring revision ACL surgery or a positive pivot shift or asymmetrical pivot shift greater than other contralateral sides
 - Secondary: patient-reported, radiological, and clinical examination findings are listed

Fact Box 7.1: FINER Criteria for a Good Research Question

F	Feasible	<ul style="list-style-type: none"> • Adequate number of subjects • Adequate technical expertise • Affordable in time and money • Manageable in scope
I	Interesting	<ul style="list-style-type: none"> • Getting the answer intrigues investigator, peers, and community
N	Novel	<ul style="list-style-type: none"> • Confirms, refutes, or extends previous findings
E	Ethical	<ul style="list-style-type: none"> • Amenable to a study that institutional review board will approve
R	Relevant	<ul style="list-style-type: none"> • To scientific knowledge • To clinical and health policy • To future research

Reproduced with permission from: Farrugia, P., et al., *Research questions, hypotheses and objectives*. Canadian Journal of Surgery, 2010. **53**(4): p. 278

Fact Box 7.2: PICOT Criteria

P	Patient/ population	<ul style="list-style-type: none"> • What specific patient population are you interested in?
I	Intervention	<ul style="list-style-type: none"> • What is your investigational intervention?
C	Comparison	<ul style="list-style-type: none"> • What is the main alternative to compare to the intervention?
O	Outcome	<ul style="list-style-type: none"> • What do you intend to accomplish, measure, improve, or affect?
T	Time	<ul style="list-style-type: none"> • What is the appropriate follow-up time to assess the outcome?

Reproduced with permission from: Farrugia, P., et al., *Research questions, hypotheses and objectives*. Canadian Journal of Surgery, 2010. **53**(4): p. 278

Take-Home Messages

- A structured and answerable research question is an essential prerequisite to the development of a research project.
- Confirmation of a knowledge gap is the initial step in the process from idea to research question and requires reviewing the available literature and ongoing research.
- The PICO approach provides a basic structure for a well-formed clinical question.
- Attention to detail and clear definitions are required for each element of the PICO model.
- Time spent developing a well-considered research question before commencing the research trial can prevent problems down the track.

7.7 Resources/Websites

<https://clinicaltrials.gov>—a database of privately and publicly funded clinical studies conducted around the world.

References

1. Ardern CL, Taylor NF, Feller JA, Webster KE. Return-to-sport outcomes at 2 to 7 years after anterior cruciate ligament reconstruction surgery. *Am J Sports Med.* 2012;40(1):41–8.
2. Benjaminse A, Gokeler A, van der Schans CP. Clinical diagnosis of an anterior cruciate ligament rupture: a meta-analysis. *J Orthop Sports Phys Ther.* 2006;36(5):267–88.
3. Cancienne JM, Brockmeier SF, Carson EW, Werner BC. Risk factors for infection after shoulder arthroscopy in a large medicare population. *Am J Sports Med.* 2018;46(4):809–14. 0363546517749212.
4. Chalmers PN, Salazar DH, Steger-May K, Chamberlain AM, Stobbs-Cucchi G, Yamaguchi K, et al. Radiographic progression of arthritic changes in shoulders with degenerative rotator cuff tears. *J Shoulder Elb Surg.* 2016;25(11):1749–55.
5. Farrugia P, Petrisor BA, Farrokhyar F, Bhandari M. Research questions, hypotheses and objectives. *Can J Surg.* 2010;53(4):278.
6. Health UNIo. *ClinicalTrials.gov.* 2012.
7. Hewison CE. STABILITY Study: a multicentre RCT comparing ACL reconstruction with and without lateral extra-articular tenodesis for individuals at high risk of graft failure. *Electronic Thesis and Dissertation Repository,* 3199. 2015. <https://ir.lib.uwo.ca/etd/3199>.
8. Hewison CE, Tran MN, Kaniki N, Remtulla A, Bryant D, Getgood AM. Lateral extra-articular tenodesis reduces rotational laxity when combined with anterior cruciate ligament reconstruction: a systematic review of the literature. *Arthroscopy.* 2015;31(10):2022–34.
9. Hoffmann T, Bennett S, Del Mar C. Evidence-based practice across the health professions-e-Book. Amsterdam: Elsevier Health Sciences; 2013.
10. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. *Designing clinical research.* Philadelphia: Lippincott Williams & Wilkins; 2013.
11. Kaura A. *Crash course evidence-based medicine: reading and writing medical papers-e-Book.* Amsterdam: Elsevier Health Sciences; 2013.
12. Kim JY, Park JS, Rhee YG. Can preoperative magnetic resonance imaging predict the reparability of massive rotator cuff tears? *Am J Sports Med.* 2017;45(7):1654–63.
13. Lee O-S, Lee YS. Diagnostic value of computed tomography and risk factors for lateral hinge fracture in the open wedge high tibial osteotomy. *Arthroscopy.* 2018, 34(4):1032–43.
14. Multicenter randomized clinical trial comparing anterior cruciate ligament reconstruction with and without lateral extra-articular tenodesis in individuals who are at high risk of graft failure 2018. <https://clinicaltrials.gov/ct2/show/NCT02018354>.
15. Oh JH, Seo HJ, Lee Y-H, Choi H-Y, Joung HY, Kim SH. Do selective COX-2 inhibitors affect pain control and healing after arthroscopic rotator cuff repair? A preliminary study. *Am J Sports Med.* 2018;46(3):679–86. 0363546517744219.
16. Rios LP, Ye C, Thabane L. Association between framing of the research question using the PICOT format and reporting quality of randomized controlled trials. *BMC Med Res Methodol.* 2010;10(1):11.
17. Sackett DL, Rosenberg WM, Gray JM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ.* 1996;312(7023):71–2.
18. Sanders TL, Kremers HM, Bryan AJ, Fruth KM, Larson DR, Pareek A, et al. Is anterior cruciate ligament reconstruction effective in preventing secondary meniscal tears and osteoarthritis? *Am J Sports Med.* 2016;44(7):1699–707.
19. Thabane L, Thomas T, Ye C, Paul J. Posing the research question: not so simple. *Can J Anesth.* 2009; 56(1):71.
20. Von Porat A, Roos E, Roos H. High prevalence of osteoarthritis 14 years after an anterior cruciate ligament tear in male soccer players: a study of radiographic and patient relevant outcomes. *Ann Rheum Dis.* 2004;63(3):269–73.
21. Willits K, Amendola A, Bryant D, Mohtadi NG, Giffin JR, Fowler P, et al. Operative versus nonoperative treatment of acute Achilles tendon ruptures: a multicenter randomized trial using accelerated functional rehabilitation. *JBJS.* 2010;92(17):2767–75.



How to Write a Study Protocol

8

Lukas B. Moser and Michael T. Hirschmann

8.1 Introduction

This first step of getting the research idea organized may be difficult for the beginner. However, it is also an important step. In the further course of the research, the young scientist will find out that writing a brief, concise, and comprehensive study protocol will save time and problems in the course of the study. If done properly it is definitely one of the most rewarding parts of the research project as it facilitates later execution of the study and the subsequent writing process (see Chapter 11.9).

Many junior researchers have difficulties starting their first research project. This might be due to a number of reasons.

Firstly, most of the scientific project ideas evolve from problems more senior orthopedic surgeons encounter during their daily practice. The seniors pass the study idea over to the younger researchers, who often do not fully understand it at first sight as they do not have the necessary broad and deep orthopedic knowledge. Hence, good mentorship from senior orthopedic researchers is of utmost importance and will

facilitate the first steps into the unknown territory of research [4].

Secondly, in many countries young orthopedic trainees may not be prepared with a basic scientific skillset at medical school. Specific courses may be rare and often do not focus enough on the challenges of junior researchers [3].

Thirdly, from a young researcher's perspective, it appears to be a waste of time to spend their valuable time on preparation of a study protocol. It is the time of publish or perish in which the young researcher is driven by the ultimate goal to get the study done, to present at a conference and finally get it published.

Fourthly, there is always a next conference and submission deadline to meet. This might mislead young researchers to gather overhasty data in order to write their abstract in time. Please do not feel rushed by upcoming deadlines. This inevitably decreases the quality of the scientific work, and the abstract might even get rejected [3].

Clearly, there is no alternative to writing a proper study protocol. Therefore, writing a good study protocol is time well spent facilitating your further study. It is also mandatory for ethical committee or grant application.

L. B. Moser · M. T. Hirschmann (✉)
Department of Orthopaedic Surgery and
Traumatology, Kantonsspital Baselland
(Bruderholz, Liestal, Laufen),
Bruderholz, Switzerland

University of Basel, Basel, Switzerland
e-mail: michael.hirschmann@ksbl.ch;
michael.hirschmann@unibas.ch

8.2 What Is a Study Protocol?

A study protocol is the central documentation of the research project including scientific, ethical, and regulatory considerations. It should provide

detailed information on planning and execution of the research project. The study protocol serves as a comprehensive guide and also represents the main document for external evaluation of the study (e.g., ethical committee, grant authorities).

However, the purpose of the study protocol is to give a concise description of the study idea, plan, and further analysis. The writing style should be brief and concise. It needs to be easy to understand even for laypersons without a medical background.

Think of a study protocol as a recipe, which should enable the cook/reader to prepare an identical meal [9].

As a study is often the result of an interdisciplinary collaboration and involves a variety of different professions such as clinicians, scientists, or statisticians, all groups should be involved at this early stage. Significant input from all groups involved is necessary.

A well-written study protocol is necessary for:

- Application for ethical approval at the local ethical committee.
- Application for a scientific grant at grant authority.
- It helps to structure and define the study idea in the discussion with the collaborators.
- It prompts rethinking the study plan and reveals possible problems and obstacles at an early stage.
- Clearly defines the responsibility of each researcher involved.
- Defines a budget and funding plan for the study.
- Defines a clear timeline for each step of the study.
- Helps to monitor the milestones and study progress.
- Facilitates writing of the later scientific article [14].

Every research protocol follows a similar basic structure. However, it is important to know that the structure of a research protocol might vary from one to another study. This is due to the fact that study protocols written for ethical committee approval might be slightly different than protocols written for grant authorities. It is also mandatory to follow exactly the instructions given by the grant authority or local ethical committee [9].

Generally, the study protocol can be divided into the following parts:

- Title page
- Background
- Study objectives (aims)
- Hypothesis
- Material and methods
 - Study design
 - Subjects
 - Sample size
 - Study procedures
 - Outcome instruments
 - Data collection
- Data management
- Statistical analysis
- Ethical considerations
- Time points and timeline
- Conflict of interest
- Insurance
- Reference section
- Annexes

To help getting started, the researcher should consider the following questions to be answered:

- What is the clinical problem to be addressed by this study?
- What is already known about the problem and topic?
- What is the study design in the study?
- What are the inclusion and exclusion criteria?
- What study subjects are investigated?
- What are the outcome instruments chosen?
- What are the primary or secondary endpoints of the study?
- What are the interventions?
- What is the experimental setup?

- How is data collection organized?
- How is the data processed and analyzed?
- What are the statistical methods?
- Are there any ethical problems to be considered?
- What is the timeline?
- What milestones are set?
- How is funding organized [22]?

Having specific answers to the aforementioned questions helps to fill every part of the study protocol. Hence, we can move on and get it started step by step.

8.3 Title Page

The title page contains the most important information at a single glance. Here, the researcher should give the full title of the study along with the names and affiliations of all people involved [4].

The project title needs to be wisely chosen as it should closely reflect the content of the study. Keep it brief and concise. The title should not pose a question, but summarize the main objective including the study type such as “randomized controlled trial.” It might also mention the subjects to be investigated [14].

The affiliations of each study team member need to be complete. For every member of the study team, the contact information including e-mail address and telephone and fax number should be given here.

If applicable state the study sponsor, in the case the study is supported by a company or grant authority [4].

8.4 Background

The background of the topic should be clearly described. Meticulously guide the reader into the topic while avoiding irrelevant information. Do not spend more than two pages on the background information. As a rule of thumb, also limit the scientific articles for reference to less than 20–30. Only describe the most important ones here. Always ask if this refer-

ence is really needed to lead the reader to the purpose of the study. It also helps to stay focused [14].

It is of utmost importance to clarify why the researcher is conducting the study. The researcher should make the reader understand why the research project is planned and what the original study idea is. As this requires a good understanding of the current knowledge in this field, a systematic review of the current literature should be performed. It facilitates the summary of current evidence and allows to put the study idea into the broader scientific context [16]. The red thread of the background section is to start from the eagle’s perspective and come closer to the topic by every sentence written [9].

Defining the research question is definitely the heart of the protocol. The research question proposed should be able to fill an existing gap of knowledge. Ideally, it should represent the logical consequence of the background explained. The question asked here should be as precise as possible [16].

This part decides the fate of the study. Clearly, if the researcher is not able to give a proper study question and explain the hypothesis here, better reconsider and do not waste the personal resources and time for this study [14] (Tables 8.1 and 8.2).

A major pitfall of study protocols is that many do not have a single study question and others just have too many. There might be several study

Table 8.1 The mnemonic FINER will help the researcher with identification of a good research question and study idea [7]

F	Feasible in terms of recruitment, expertise, funding, resources, and scope
I	Interesting to the scientific community
N	Novel, providing new findings and confirming or rejecting previous findings
E	Ethical approval should be possible
R	Relevant to scientists, clinicians, and future research

Table 8.2 The PICO acronym helps to formulate a testable question [15]

P	Patient, problem
I	Intervention or exposure
C	Comparison
O	Outcome

questions of interest with regard to the research project that need to be answered. However, the researcher should stay focused on the most important research question. Secondary research questions are often of explorative nature, because sample size calculation is based on the primary research question. Do not overload the project with too many research questions. The researcher will then lose the red thread and the reader will be confused. The researcher also has to consider that each research question requires a hypothesis. Defining the research questions and objectives in advance helps the researcher to avoid reporting significant results more likely than negative results (outcome reporting bias) [16].

8.5 Study Objectives (Aims)

After performing the literature review with a focus on the research question, the researcher needs to define the objectives, outcomes, and hypothesis. A clear definition of these factors is recommended [9] (Table 8.3).

Choose the study objectives wisely and restrictively. If the researcher decides to use more objectives, (s)he should distinguish between primary and secondary objectives or outcomes. Altogether, do not choose more than three to five objectives [4].

Practical Case Example

A clinical study investigating if the femoral prosthetic design in total knee arthroplasty (TKA) influences the patellar loading and functional scores in TKA with unresurfaced patellae

The primary purpose of the study is to investigate if the femoral shape of the different TKA models, (group P) and (group A), influences the BTU distribution pattern at SPECT-CT in TKA with unresurfaced patellae. Since femoral TKA malrotation, patellar height, and TT-TG are

known to be possible causes of patellar overstuffing inducing a higher patellar BTU, these possible bias parameters will be calculated and compared between the two groups.

The secondary purpose is to compare postoperative functional results measured at 1 and 2 years postoperative.

An exact definition of outcomes is mandatory to standardize the outcome measures. Note the time of assessment and the unit of measures. If previous definitions or validations are preferred, do not forget to quote the references. If the researcher compares outcomes, (s)he needs to state an overall goal of the comparison in the protocol. The reader should understand if the goal was to show superiority, equivalence, or non-inferiority [16].

Sometimes, outcomes are not easy to determine. Some outcomes might be subjective (e.g., pain); others are more objectively measurable (e.g., range of motion).

Surrogate endpoints are alternative endpoints that are faster to assess than long-term clinical endpoints [8]. The effect of the intervention on the surrogate endpoint has to correspond to the effect on clinical outcome. However, this effect is often difficult to predict, and therefore surrogate endpoints should only carefully be used [12].

Furthermore, the researcher needs to identify all potential confounders. Confounders are defined as additional factors that distort the effect of the treatment or exposure on a predicted outcome. However, they have an impact on both, the exposure and outcome, while not being part of the causal pathway. Confounding factors might lead to the situation that falsely a correlation between treatment and outcome might be seen or a true relationship is hidden. Therefore, confounders should be carefully considered. Ideally, confounders are equally distributed in randomized controlled studies. Be aware that this is not the case in observational data [18].

Table 8.3 It helps when the researcher defines the objectives of the study with regard to the SMART criteria [6]

S	Specific
M	Measurable
A	Achievable
R	Realistic
T	Time-related

Clinical Vignette

Age being a confounder in a study investigating the association of physical activity and knee pain

Younger individuals tend to be more active and have a lower risk of knee pain. If the proportion of young people being more active with less pain is not equally distributed, the association between activity and knee pain might be overestimated [21].

Vavken et al. investigated the consideration of confounders in 126 controlled studies published in journals with a high-impact factor. Although 16% of the studies discussed confounding factors, they did not adjust their subsequent analysis. Only 1/3 of the authors controlled for confounding factors [21]. However, confounders need to be identified in the study protocol. Only then the later analysis can be adjusted [2].

8.6 Hypothesis

The hypothesis is based on the supposed relation between variables. State the hypothesis in the null form. The null hypothesis says that there is no relationship between variables and the researcher is going to challenge that statement. It represents the contradictory of what to expect. Statistical testing will reveal the probability that the null hypothesis is true or false. The alternative hypothesis is recognized when the null hypothesis is rejected. It represents the outcome the researcher would expect [14].

8.7 Material and Methods

The methods section needs to explain how the hypothesis is tested. It is of utmost importance to find the optimal methodological approach for the study. After a first draft, the researcher often has to critically revise the methodology considering input from all collaborators [9]. An exact and precise description of the methods used is the most important part of the study protocol. At the beginning state the study design and then continue with the description of the study subjects. The reader must be able to understand how the researcher

identified the subjects or patients, what were the inclusion and exclusion criteria (study population), and what is investigated (study procedure). Explain each step of the study procedure meticulously. Only then an independent reader is able to follow and reproduce the study [4].

The methods should include the following aspects: design and setting, subjects, sample size, description of the study procedure, data collection, data management, and finally the statistical tests to be used.

8.8 Study Design

Precisely define the study type and design. One can differentiate between research on primary and secondary data. Whereas research on primary data means the researcher is performing the actual study (e.g., clinical, experimental, and epidemiological studies), research on secondary data describes the process of analyzing studies that have already been performed (e.g., systematic review, meta-analysis).

Chapter 11.9 provides a comprehensive list of all possible study types. Although in practice often difficult, the researcher should always aim for the study type providing the best scientific evidence and quality level.

8.9 Subjects

The study subjects or patients included need to be characterized in detail. Including a flow chart showing the recruitment, screening, inclusion, and exclusion criteria will help the reader to understand this process (Fig. 8.1).

The eligibility criteria need to be mentioned along with all inclusion and exclusion criteria. Inclusion criteria determine which subjects or patients are going to be included into the study. All other criteria limiting eligibility are considered as exclusion criteria [19].

Explain how and why the researcher decided to choose the sample size used in the study. This

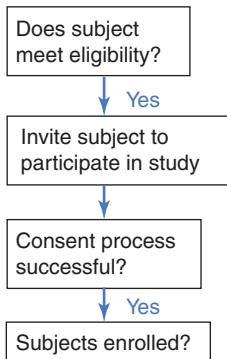


Fig. 8.1 Flowchart showing enrollment process

needs to be based on a sample size calculation. Keep in mind that an inclusion of vulnerable population (children, cognitively impaired adults) requires special justification [5].

Provide a detailed and unambiguous list of the eligibility criteria as in the example below:

Enclose a statement which states that the participation of the patient is entirely voluntary, and an unexplained withdrawal is possible at any time having no impact on patient management [16].

Practical Case Example

Any subject is entitled to withdraw from this clinical investigation for any reason without obligation or prejudice to further treatment.

Declare how the patient consented to participate in the study.

Practical Case Example

The patient shall give his written consent to participate in the clinical documentation. The signed consent form remains within the clinical study case file.

8.10 Sample Size

Choosing the appropriate study sample is paramount for successful research. The study sample ensures the sufficient power of the study and might allow conclusive inferences [16]. Consider that having a small sample size might not be powerful enough to prove a difference, although true difference actually exists. However, a too large sample size might be a waste of resources [14].

8.11 Study Procedures

In this section, the researcher should clearly describe what is going to be done with the subjects, how, and when. Provide a detailed description of the study procedure with the planned interventions.

8.12 Data Collection

Describe clearly who and how to obtain the data, how to record it, and how long it will be stored [4]. Name the locations where the data is going to be collected, relevant events regarding recruitments, exposer, or follow-up [20].

State the data-collecting instruments (e.g., electronic forms or paper forms). Explain the validation procedure for the instruments used. For outcome instruments, provide original references. If the researcher is going to develop a novel type of data-collecting instruments, as a checklist or a questionnaire, the researcher needs to provide all information. Do not forget to annex the original document to the study protocol.

Practical Case Example

This is a retrospective case series study. No additional examinations other than the clinical routine will be conducted. All data is stored in a centralized electronic database at the orthopedic research unit (...). The data has been already collected in our ethical committee approval registry.

8.13 Data Management

The researcher needs to make a comprehensive statement on the data management including data entry and monitoring. Describe how the researcher plans to ensure privacy protection (e.g., anonymization) and how long the data is stored. Declare who is able to access the data, and state that all involved people underline medical confidentiality. Add the statement by the IEC and regulatory authorities for permitting data access, if applicable [16].

Mention how to introduce the data into the database. Describe what programs the researcher will use and how to perform data analysis. The best approach is to make a data analysis plan

where the researcher describes which measurements were done with which variable and which statistical tests to apply. Also, record how to deal with missing data [9].

8.14 Statistical Analysis

A researcher should have basic knowledge of statistics [11]. However, consulting a professional statistician might help to improve the quality of the research methodology. At best the statistician is already involved in the process of study planning, as the researcher needs to translate the research question into a statistical problem. All statistical relevant information must be included in the sufficient level of details: exploratory or descriptive statistics, level of significance, type of outcome, effect measures including confidence intervals, type of sample (unpaired or paired), data distribution (not normally or normally distributed data), and the statistical software used [16].

Just naming the tests that were used is not enough. The researcher needs to state why a test was chosen and which test for which purpose [4].

Practical Case Example

A clinical study investigating if the femoral prosthetic design in total knee arthroplasty (TKA) influences the patellar loading and functional scores in TKA with unresurfaced patellae

Descriptive statistics (means, medians, quartiles, ranges, standard deviations) will be performed to assess the demographics of the patient population. Alignment and TKA component position in all planes (sagittal, coronal, and rotational) are noted in degree. Mean and absolute relative BTU of corresponding quadrants of the two groups will be compared with a *t*-test.

Nonparametric Spearman's correlation coefficients will be used to correlate the patella height, the lateral tilt, and the components' alignment measurements with the intensity of the tracer uptake in each area of interest. Postoperative KSS scores of the two groups at 1 and 2 years postoperative will be compared with a *t*-test.

All data will be analyzed by an independent professional statistician.

8.15 Ethical Considerations

This part is particularly important for the ethical committee. It will be carefully read; in particular the ethical risks will be discussed between the members of the ethical board. Present a risk-benefit assessment providing information on potential benefits, risks, or inconveniences for the individual study subjects. If necessary, justify the inclusion of vulnerable populations. Include a statement that the researcher conducts the study according to the study protocol and good clinical practice (GCP) [16].

Keep in mind that any changes of the study protocol need to be amended to the ethical committee and the regulatory authorities.

Practical Case Example

The study will be performed in accordance with the declaration of Helsinki and the directives for good clinical practice (GCP) standards. Ethical approval will be obtained from local ethical committee in accordance with national law.

8.16 Time Points and Timeline

Creating a table including the time points and timeline is useful for several reasons. It forces the researcher to set a time frame and think about important deadlines and targets. The reader understands the sequence of events at a single glance, and the timetable shows the reviewer that the project is expected to be completed in the foreseeable future [13] (Tables 8.4 and 8.5).

Table 8.4 Practical case example of time points

Writing the study protocol	January
Data collection	March–August
Data analysis	September
Discussion of results among research staff	October
Writing of scientific article	November
Submission to peer-reviewed journal	December

Table 8.5 Practical case example of a timeline

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Writing the study protocol												
Data collection												
Data analysis												
Discussion of results among research staff												
Writing of scientific article												
Submission to peer-reviewed journal												

8.17 Conflict of Interest

Do not forget to disclose the conflicts of interests, such as nonfinancial or financial relationships with the industry. If applicable, state the study sponsor. Declare how the sponsor contributed and what benefits the researcher will obtain by this agreement [4].

These disclosures facilitate the identification of possible sponsor bias. Most journals require a disclosure form of all contributing authors. A short statement of relevant information will be listed on the paper in abbreviated form [17].

Practical Case Example 1

One of the authors (XY) is a consultant for (company).

Practical Case Example 2

The study was supported by a financial grant from (company). The external funding source did not have any influence on the investigation.

Practical Case Example 3

The authors of this paper declare no conflict of interest.

8.18 Insurance

The sponsor of the study is obliged to provide clinical trial insurance. If the researcher is conducting a clinical study, (s)he needs to compen-

sate serious adverse events (as injuries) the patients suffer during their participation. However, rules and requirements of trial insurance vary from country to country. Therefore, the researcher needs to check the local insurance policy [10].

Practical Case Example 1

The sponsor will secure and maintain in full force and effect, throughout the duration of the clinical investigation, a clinical trial insurance which was required in line with national regulations and which will be evidenced by an insurance certificate.

Practical Case Example 2

No subject insurance is required as it is a retrospective study and the patient is not put at any risk or has to show up for follow-up.

8.19 Reference Section

The reference section should be organized as in a scientific paper (see Chapter 11.9). Provide original references, and number them consecutively or in the order requested. For every statement made, a citation should be given. It is important that only articles are cited which really add to the study protocol. There should not be more than 30–40 references in the reference list [4].

8.20 Annexes

In particular for major grant applications, annexes could be added at the end. These might include informed consent sheet, study questionnaire or CRF, approval of ethical committee if already obtained, and CV of the principal and co-investigator [1].

Take-Home Message

- Writing a brief and comprehensive study protocol is always the first step of your research project.
- It facilitates later execution of your study and the subsequent writing process.
- Make sure that your study protocol includes all relevant information on your study idea, plan, and further analysis.
- Furthermore, as a comprehensive guide, it represents the main document for external evaluation of your study.
- The idea of the study protocol is to convey to any individual the purpose of the study and the probable impact of the project.

References

1. Al-Jundi A, Sakka S. Protocol writing in clinical research. *J Clin Diagn Res.* 2016;10:ZE10–3.
2. Babyak MA. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosom Med.* 2004;66:411–21.
3. Bando K, Sato T. Did you write a protocol before starting your project? *Gen Thorac Cardiovasc Surg.* 2015;63:71–7.
4. Barletta JF. Conducting a successful residency research project. *Am J Pharm Educ.* 2008;72:92.
5. Brody BA, McCullough LB, Sharp RR. Consensus and controversy in clinical research ethics. *JAMA.* 2005;294:1411–4.
6. Doran G. There's a S.M.A.R.T. way to write management's goals and objectives. *Manag Rev.* 1981;70:35–6.
7. Doran PC. Understanding dentures: instructions to patients. *J Colo Dent Assoc.* 1981;59:4–8.
8. Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? *Ann Intern Med.* 1996;125:605–13.
9. Fronteira I. How to design a (good) epidemiological observational study: epidemiological research protocol at a glance. *Acta Medica Port.* 2013;26:731–6.
10. Ghooi RB, Divekar D. Insurance in clinical research. *Perspect Clin Res.* 2014;5:145–50.
11. Gilmore SJ. Evaluating statistics in clinical trials: making the unintelligible intelligible. *Australas J Dermatol.* 2008;49:177–84; quiz 185–176.
12. Huang Y, Gilbert PB. Comparing biomarkers as principal surrogate endpoints. *Biometrics.* 2011;67:1442–51.
13. Kanji S. Turning your research idea into a proposal worth funding. *Can J Hosp Pharm.* 2015;68:458–64.
14. O'Brien K, Wright J. How to write a protocol. *J Orthod.* 2002;29:58–61.
15. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club.* 1995;123:A12–3.
16. Rosenthal R, Schafer J, Briel M, Bucher HC, Oertli D, Dell-Kuster S. How to write a surgical clinical research protocol: literature review and practical guide. *Am J Surg.* 2014;207:299–312.
17. Sade RM, Akins CW, Weisel RD. Managing conflicts of interest. *J Thorac Cardiovasc Surg.* 2015;149:971–2.
18. Skelly AC, Dettori JR, Brodt ED. Assessing bias: the importance of considering confounding. *Evid Based Spine Care J.* 2012;3:9–12.
19. Van Spall HG, Toren A, Kiss A, Fowler RA. Eligibility criteria of randomized controlled trials published in high-impact general medical journals: a systematic sampling review. *JAMA.* 2007;297:1233–40.
20. Vandembroucke JP, von Elm E, Altman DG, Gotzsche PC, Mulrow CD, Pocock SJ, et al. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation and elaboration. *PLoS Med.* 2007;4:e297.
21. Vavken P, Culen G, Dorotka R. Management of confounding in controlled orthopaedic trials: a cross-sectional study. *Clin Orthop Relat Res.* 2008;466:985–9.
22. Warren MD. Aide-memoire for preparing a protocol. *Br Med J.* 1978;1:1195–6.

The Ethical Approval Process

9

Karren Takamura and Frank Petrigliano

9.1 Important Documents in Biomedical Ethics

Prior to the Nuremberg Code, issued by the Nuremberg Military Tribunal during the Nuremberg trials in 1947 (also known as the “Doctors’ trial”), there was no generally accepted code of ethics for human research. The Nuremberg Code is a ten-point statement of ethics to prevent abuse of human research subjects and establishes that participation in research must be voluntary (Fact Box 9.1). While the Nuremberg Code were never formally adopted by any state or international agency, it is considered one of the most influential documents in human medical research and served as the basis for documents that later followed [5].

Fact Box 9.1: Nuremberg Code

1. The voluntary consent of the human subject is absolutely essential. This means that the person involved should have legal capacity to give consent; should be so situated as to be able to exercise free power of choice, without

the intervention of any element of force, fraud, deceit, duress, overreaching, or other ulterior form of constraint or coercion; and should have sufficient knowledge and comprehension of the elements of the subject matter involved, as to enable him to make an understanding and enlightened decision. This latter element requires that, before the acceptance of an affirmative decision by the experimental subject, there should be made known to him the nature, duration, and purpose of the experiment; the method and means by which it is to be conducted; all inconveniences and hazards reasonably to be expected; and the effects upon his health or person, which may possibly come from his participation in the experiment. The duty and responsibility for ascertaining the quality of the consent rests upon each individual who initiates, directs or engages in the experiment. It is a personal duty and responsibility which may not be delegated to another with impunity.

2. The experiment should be such as to yield fruitful results for the good of society, unprocurable by other methods or means of study, and not random and unnecessary in nature.

K. Takamura · F. Petrigliano (✉)
Department of Orthopaedic Surgery, University of California, Los Angeles, USA
e-mail: KTakamura@mednet.ucla.edu; fpetrigliano@mednet.ucla.edu

3. The experiment should be so designed and based on the results of animal experimentation and a knowledge of the natural history of the disease or other problem under study, that the anticipated results will justify the performance of the experiment.
4. The experiment should be so conducted as to avoid all unnecessary physical and mental suffering and injury.
5. No experiment should be conducted, where there is an a priori reason to believe that death or disabling injury will occur; except, perhaps, in those experiments where the experimental physicians also serve as subjects.
6. The degree of risk to be taken should never exceed that determined by the humanitarian importance of the problem to be solved by the experiment.
7. Proper preparations should be made and adequate facilities provided to protect the experimental subject against even remote possibilities of injury, disability, or death.
8. The experiment should be conducted only by scientifically qualified persons. The highest degree of skill and care should be required through all stages of the experiment of those who conduct or engage in the experiment.
9. During the course of the experiment, the human subject should be at liberty to bring the experiment to an end, if he has reached the physical or mental state, where continuation of the experiment seemed to him to be impossible.
10. During the course of the experiment, the scientist in charge must be prepared to terminate the experiment at any stage, if he has probable cause to believe, in the exercise of the good faith, superior skill and careful

judgment required of him, that a continuation of the experiment is likely to result in injury, disability, or death to the experimental subject.

Reference: reproduced from “Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10”, Vol. 2, pp. 181–182. Washington, D.C.: U.S. Government Printing Office, 1949.

The Declaration of Helsinki was originally developed by the World Medical Association (WMA) in 1964 in order to establish a set of ethical principles regarding research involving human subjects. This document is widely regarded as the cornerstone for medical research involving human subjects, including identifiable human material and data [19]. The declaration is primarily written for physicians, though it is intended as a guideline for the broader research community. The fundamental principles of the Declaration of Helsinki include respect for the individual, right to self-determination, protection of privacy, and the right to make informed decisions [19]. The document proclaims that the physician’s duty is solely to the participant; the participant’s welfare supersedes the interest and benefit of science and society [19]. This document has undergone multiple revisions, most recently in 2013, addressing issues more relevant to countries with limited resources, such as post-trial access to interventions, compensation and treatment for individuals harmed during participation in research, access to clinical trial for underrepresented groups, and need for dissemination of research results [12, 19].

9.2 History of Biomedical Human Research in the United States

In the United States, the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research was created by the

National Research Act of 1974 to identify ethical principles and guidelines for conducting human research in response to the Tuskegee syphilis experiment [5]. The Tuskegee syphilis experiment was a prospective clinical study conducted between 1932 and 1972 by US Public Health Service to study the natural history of untreated syphilis in rural African-American males in Alabama. The men were never told of their diagnosis, and although penicillin was found to be an effective cure for the disease, treatment was withheld [16]. A whistle-blower by the name of Peter Buxtun revealed the story to the press, leading to public outrage and major changes in US law and regulation on how human research is conducted [2, 16]. Other controversial research projects in the United States during this era include the Stanford prison experiment (1971), where the participants were unable to withdraw from the study [20], and Project MKUltra (1950s–1973) where subjects were not informed of their participation in the studies [14].

In 1978, the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research issued the Belmont Report, which established three core principles (Fact Box 9.2) for research involving human studies: respect for persons, beneficence, and justice [18]. The applications of these principles led to the consideration of the following requirements: informed consent, risk/benefit analysis, and patient selection [17].

Fact Box 9.2: Belmont Report, Three Core Principles for Research Involving Human Studies [18]

1. Respect for persons
2. Beneficence
3. Justice

9.3 Common Rule (Part 46, Protection of Human Subjects, of Title 45, Public Welfare, in Code of Federal Regulations)

The National Research Act of 1974 established a set of guidelines for research involving human subjects, introducing the concept of the

Institutional Review Board (IRB). The basic provisions of the IRB are outlined by the Federal Policy for the Protection of Human Subjects, or “Common Rule” was published and codified by 15 federal department and agencies [3]. The Common Rule also outlines the basic provisions for IRB, informed consent, and assurances of compliance [3]. Any research that is conducted by or for these federal agencies must abide by the “basic policy for protection of human research subjects” (also known as Subpart A, Part 46, Protection of Human Subjects, of Title 45, Public Welfare, in Code of Federal Regulations (46 CFR 45)).

The IRB is an independent, administrative group tasked with the responsibility of reviewing and approving research on human subjects, with the purpose of protecting the rights and welfare of human subjects. IRBs, and human subject research in general, are regulated by the Office for Human Research Protections (OHRP), an organization within the Department of Health and Human Services (HHS). The goal of the IRB is to insure that proper informed consent is obtained and documented, risks to subjects are minimized, research design is sound and do not unnecessarily expose subjects to risk, patient selection is equitable, appropriate data monitoring provisions are in place, and privacy and confidentiality of the subjects are maintained [3]. The IRB also has the power to terminate or suspend any research that is not in accordance with the policy [3]. The IRB is comprised of scientists, lay community members, physicians, and lawyers [3]. The average size of the IRB is 14 members. One study found internal medicine to be most commonly represented and orthopedic surgery to be the least represented, among physician members of IRBs [10].

The Common Rule also includes regulations for addressing vulnerable populations, such as pregnant women, fetuses, in vitro fertilization, prisoners, and children. Subpart B of 46 CFR 45 affords special protections to pregnant women, fetuses, and neonates of uncertain viability or nonviable neonates. Research must directly benefit the mother and/or the fetus; if not, risks to the fetus must be minimal, and the purpose of the study must be “the development of important

biomedical knowledge that cannot be obtained by any other means.” Subpart C of 46 CFR 45 affords special protections to prisoners, to ensure that they are not exploited, but at the same time given equal opportunity to participate in research studies.

9.4 Research in Children and Adolescents

Children and adolescents comprise an important group of study participants in orthopedics. Importantly, this is a vulnerable population that warrants additional protections. The risks and benefits must be carefully evaluated, to prioritize the welfare of the patient, while recognizing the positive potential benefits of research. This risk/benefit evaluation often impacts participant inclusion/exclusion criteria and consideration for how the standard of care may be affected by research activities [8].

One of the major differences in terms of research in children is that, by definition, they are unable to provide informed consent [5]. Instead, children can provide assent, which is defined in the policy as “a child’s affirmative agreement to participate in research” (§46. 402 [b]), and parents or legal guardians can grant permission for their child to participate (§46. 402 [c]). The process of obtaining assent should address the developmental stage of the child and provide opportunities for the child to discuss their willingness or unwillingness to participate, the degree to which the child has control over the participation decision, and whether certain information will or will not be shared with the parents [8]. Typically, local IRBs will provide guidelines for the age and conditions where a child’s assent is required.

If the research involves acute illnesses or injuries, the investigators and IRB should provide for “ongoing process for permission and assent” to accommodate for the evolving understanding in the changes of the child’s medical and mental condition [8]. Waivers of parental permission for adolescent participation should be considered by the IRB when the “research is important to the health

and well-being of adolescents and cannot reasonably or practically be carried out without the waiver” or if the research involves treatments that adolescents can receive without parental permission (may differ by state law) [3, 8]. Additionally, the investigators also need to present evidence that the adolescents have the ability to understand the research and their rights as participants and the research protocol must contain safeguards to protect the interests of the adolescent, consistent with the risk presented to them [8].

9.5 IRB Submission and Approval Process

The definition of research involving human subjects set forth by the common law [45 CFR 46.102 (d)] states that it is “[a] systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge” [3]. A majority of research involving human subjects adheres to this definition and, subsequently, requires IRB approval, but there is a subset of research that appears to adhere to this definition that does not require IRB oversight. This includes certain quality improvement and quality assurance initiatives, case reports, and case reviews [13]. The Food and Drug Administration (FDA) provides two general guidelines indicating that a quality improvement or quality assurance initiatives constitute human research if the investigators will seek publication in a scientific/national journal or presentation at a national meeting or if the results will be applicable in a wider setting [13]. With regard to case series or case reports, there is no regulatory guidance on this particular issue [13]. If there are any questions of whether IRB approval is required for a particular study, it is advisable to consult with the IRB prior to starting the study.

The IRB submission process can be an arduous and time-consuming task that may involve multiple modifications and revisions. It can especially be cumbersome in multicenter trials involving multiple IRBs, and the variability between institutions can hinder multi-institutional

research [4]. One study in the United Kingdom found that only 24% of studies submitted were approved without modifications [11]. Common reasons for proposal rejection were improperly designed consent form, poor study design, unacceptable risk to subjects, and ethical and legal reasons (Fact Box 9.3) [10]. Some suggestions to the young investigator navigating the IRB include collaborating with an experienced mentor, familiarizing oneself with the IRB guidelines and procedures of the research site(s), and discussing the protocol with the IRB prior to submission [10].

Fact Box 9.3: Common Reasons for Proposal Rejection [10]

1. Improperly designed consent form
2. Poor study design
3. Unacceptable risk to subjects
4. Ethical and legal reasons

There are three levels of review that are identified by federal regulation: expedited review, full or convened review, and exemption from review. If the study poses only “minimal risk” to the subject, the study may be suitable for expedited review. “Minimal risk” is defined as such “that the probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examination or tests” [45 CFR 46.102(i)]. Studies that may be suitable for expedited review based on aforementioned definition are shown in Table 9.1. For minimal risk studies, there is considerable variability in the IRB process [7]. These studies are not reviewed by the full IRB and typically reviewed by a subcommittee or administratively within the office [13].

A study may be exempt from review if it meets one of six of the federally defined exempt categories. Examples include research conducted in established or commonly accepted educational settings, retrospective review of existing data, doc-

Table 9.1 OHRP expedited review categories [3]

1. Clinical studies of drugs and medical devices only if one of the following conditions are met:
(a) Research on drugs for which an investigational new drug application is not required
(b) Research on medical devices for which an investigational device exemption application is not required or the medical device is cleared and approved for marketing and is being used for which it has been cleared and approved for
2. Collection of blood samples (finger stick, heel stick, ear stick, venipuncture)
3. Prospective collection of biological samples for research purposes by noninvasive means
4. Collection of data through noninvasive measures (excluding X-rays and microwaves)
5. Research involving materials that have been collected or will be collected for non-research purposes
6. Collection of voice, video, digital, or image recordings for research
7. Research on individual or group characteristics or behavior or research involving interviews, surveys, etc.
8. Continuing review of research previously approved by IRB if:
(a) Enrollment of new subjects is closed or if all subjects have completed the research-related interventions or if the research remains active only for long-term follow-ups
(b) No subjects have been enrolled and no additional risks have been identified
(c) Remaining research activities are limited to data analysis
9. Continuing review of research (not conducted under an investigational new drug application or investigational device exemption that does not fit under items 2 through 8) for which the IRB has determined that the research involves no greater than minimal risk and no additional risks have been identified

* This concise summary is modified and abbreviated from the OHRP Expedited Review Categories [3]. Refer to the OHRP for complete information

uments, specimens, and taste and food quality evaluation (for full list, see 45 CFR 46.101 (b)). However, an exemption of a study must be made by the IRB, and no further notification is required from the IRB if that status is granted. A study requires full review if it poses “greater than minimal risk.” Examples include Phase I, II, and III clinical trials, studies involving vulnerable populations, and studies including investigational devices.

9.6 Informed Consent

The Common Rule sets forth the components of informed consent in 45 CFR 46.116 [3]:

- A statement that the study involves research, its purpose, its duration, description of procedures, and identification if the research is experimental
- Description of any “reasonably foreseeable” risks and discomforts
- Description of any possible benefits to the participant or others that may be reasonably expected from the study
- Disclosure of appropriate alternative interventions (if any) that may be advantageous to the participant
- Statement describing the extent (if any) to which confidentiality of subject data is maintained
- If the research involves more than minimal risk and explanation of any compensation or if medical treatments are available if an injury occurs
- Information of the contact person regarding questions about the research, participants’ rights, and the contact person when an injury occurs

Additionally, the IRB may request for additional elements when appropriate:

- Risks that may be “unforeseeable” (e.g., to the embryo or fetus if the participant becomes pregnant)
- Anticipated circumstances where the investigator will terminate the participant’s involvement in the study without their consent
- Additional costs that the participant may incur
- Consequences if a participant decides to withdraw from the study and procedures for “orderly termination”
- A statement that “significant new findings” which may affect the participant’s willingness to continue during the course of the study will be provided to the participants
- An approximate number of participants in the study

The informed consent process may be expedited or waived if the research is considered “minimal risk,” if the waiver or alteration of the consent does not adversely affect the participant’s rights and welfare, if the research cannot be practically achieved without the waiver/alteration, and if pertinent information will be given to the patients after the study if appropriate.

Studies have demonstrated that participants’ understanding of the informed consent is oftentimes inaccurate or incomplete [1, 9]. Additionally, one of the most requested changes required for study approval by the IRB are modifications to the consent form [10]. A systematic review found use of multimedia and enhanced consent forms had limited success in improving participants’ understanding, but having a study team member or a neutral educator spending one-on-one time with the participant was found to be the most effective way of improving their understanding [6]. It is important to keep in mind that the investigator’s obligations regarding the consent process does not end once the participant signs the form; if the investigator believes this to be true, they may be committing a “serious disservice to the participant by not observing the ethical standards” [13]. For example, the investigator or a member of the research team should be available to answer questions regarding the study at any point in the study.

Take-Home Messages

- Institutional regulations and laws exist to protect human subjects in research.
- The IRB process may be difficult, oftentimes with several modifications, but working with the IRB prior to submission is helpful and recommended.
- Obtain proper informed consent (following the requirements set forth by 45 CFR 46.116), and keep in mind that the investigator’s obligations do not end as soon as the participant provides their signature on the form.

References

1. Barrett R. Quality of informed consent: measuring understanding among participants in oncology clinical trials. *Oncol Nurs Forum*. 2005;32:751–5. <https://doi.org/10.1188/05.ONF.751-755>.
2. Corbie-Smith G. The continuing legacy of the Tuskegee Syphilis Study: considerations for clinical investigation. *Am J Med Sci*. 1999;317:5–8.
3. Department of Health and Human Services NIOH, Office for Protection for Research Risks Code of Federal Regulations - Title 45 Public Welfare CFR 46.
4. Dyrbye LN, et al. Medical education research and IRB review: an analysis and comparison of the IRB review process at six institutions. *Acad Med*. 2007;82:654–60. <https://doi.org/10.1097/ACM.0b013e318065be1e>.
5. Fischer BA IV. A summary of important documents in the field of research ethics. *Schizophr Bull*. 2006;32:69–80. <https://doi.org/10.1093/schbul/sbj005>.
6. Flory J, Emanuel E. Interventions to improve research participants' understanding in informed consent for research: a systematic review. *JAMA*. 2004;292:1593–601. <https://doi.org/10.1001/jama.292.13.1593>.
7. Hirshon JM, Krugman SD, Witting MD, Furuno JP, Limcangco MR, Perisse AR, Rasch EK. Variability in institutional review board assessment of minimal-risk research. *Acad Emerg Med*. 2002;9:1417–20.
8. Institute of Medicine BoHSP, Committee on Clinical Research Involving Children. Ethical conduct of clinical research involving children. Washington, DC: National Academies Press; 2004.
9. Joffe S, Cook EF, Cleary PD, Clark JW, Weeks JC. Quality of informed consent in cancer clinical trials: a cross-sectional survey. *Lancet*. 2001;358:1772–7. [https://doi.org/10.1016/S0140-6736\(01\)06805-2](https://doi.org/10.1016/S0140-6736(01)06805-2).
10. Jones JS, White LJ, Pool LC, Dougherty JM. Structure and practice of institutional review boards in the United States. *Acad Emerg Med*. 1996;3:804–9.
11. Kent G. Responses by four Local Research Ethics Committees to submitted proposals. *J Med Ethics*. 1999;25:274–7.
12. Ndebele P. The Declaration of Helsinki, 50 years later. *JAMA*. 2013;310:2145–6. <https://doi.org/10.1001/jama.2013.281316>.
13. Parvizi J, Tarity TD, Conner K, Smith JB. Institutional review board approval: why it matters. *J Bone Joint Surg Am*. 2007;89:418–26. <https://doi.org/10.2106/JBJS.F.00362>.
14. Project MK Ultra, the Central Intelligence Agency's' Program of Research into Behavioral Modification (1977), First Session ed. Washington: U.S. Government Printing Office (copy hosted at the New York Times website); 1977.
15. Trials of War Criminals before the Nuremberg Military Tribunals under Control Council Law No. 10, Vol. 2. Washington, D.C.: U.S. Government Printing Office; 1949. p. 181–2.
16. Reverby SM. Examining Tuskegee: the infamous syphilis study and its legacy. Chapel Hill: The University of North Carolina Press; 2009.
17. United States. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. The Belmont report: ethical principles and guidelines for the protection of human subjects of research. DHEW Publication no (OS) 78-0012. The Commission; for sale by the Supt. of Docs. Bethesda: U.S. Government Printing Office; 1978.
18. United States. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research., United States. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. Appendix, the Belmont report: ethical principles and guidelines for the protection of human subjects of research. DHEW publication no (OS) 78-0013. The commission; for sale by the Supt. of Docs. Bethesda: U.S. Government Printing Office; 1978.
19. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310:2191–4. <https://doi.org/10.1001/jama.2013.281053>.
20. Zimbardo PG. On the ethics of intervention in human psychological research: with special reference to the Stanford prison experiment. In: Pimple KD, editor. *Research ethics*. New York: Routledge; 2016.



How to Assess Patient's Outcome?

10

Yuichi Hoshino and Alfonso Barnechea

10.1 Introduction

In orthopedic surgery, we find ourselves in the need to be able to report the results of whatever procedure we do, whether within the context of a specific pathology or in a broader sense when analyzing a given surgical technique. Clearly, clinical results after any kind of treatment, whether non-operative or operative, which might be expressed through specific criteria, are defined as an outcome. For example, possible outcomes to be measured after arthroscopic Bankart repair could be the postoperative, clinician-measured range of motion or the ability of the patient to resume his/her previous sports level.

In probability theory, an outcome is a pre-defined result of a given event. An outcome is also understood as what is ultimately achieved by an activity. An outcome is useful for the determination and evaluation of the results of what we do. For example, the event is our intervention (e.g., an arthroscopic procedure), and the outcome would be the measured result (e.g., the level of return to sports a given time after the procedure).

Y. Hoshino (✉)

Department of Orthopaedic Surgery, Graduate School of Medicine, Kobe University, Kobe, Japan

A. Barnechea

Department of Orthopedic Surgery, Edgardo Rebagliati Martins National Hospital, Lima, Peru

Note that one intervention or event can have multiple outcomes, but, within an outcome, the results are mutually exclusive (one just cannot have at the same time a grade-II laxity and grade-III laxity 3 months after primary ACL reconstruction).

Outcomes as such need to be predefined, that is, if one wants to report a result, one needs to know how to define it in terms that are both understandable by peers and reproducible. In addition, an outcome should be measurable. Typically, an outcome instrument is a measurable scale, score, or rating, which allows assessment of outcomes in a standardized manner and comparison within the patients group or with others. Outcome instruments allow a reliable and reproducible documentation of patient's health or function at the same or different time points.

10.2 What Type of Outcome Instruments Exist and What Are They Used for?

Outcome instruments can be quantitative or qualitative. (Quantitative instruments measure outcomes with numeric values, which can be continuous (e.g., the measurement of tibial displacement in millimeters on a KT-1000 test or the number of degrees of passive shoulder external rotation) or discontinuous (the number of stairs a person can climb). Qualitative instruments measure more subjective results, such as the tenderness of a

joint on palpation or on motion or the overall degree of satisfaction after a procedure).

A further differentiation can be made based on their area of evaluation. One can differentiate outcome instruments assessing health status more globally, disease specific, or specifically each joint, organ, or region.

Another way of separating outcome instruments is based on their source of reporting. The clinician- or independent observer-based instruments can be distinguished from patient-reported instruments.

Clinician-reported outcome instruments: Within these outcome instruments, outcomes are measured objectively by someone other than the patient, such as a physician, technician, or researcher. Typical examples are:

- Anterior tibial displacement in millimeters measured by KT-1000 testing
- Degrees of shoulder abduction by manual examination
- Degrees of knee varus by X-ray measurements

Patient-reported outcome instruments: These outcome instruments are subjective and are reported by the patient such as:

- Degree of shoulder pain at rest measured by the visual analogue scale
- Ability to perform activities such as opening a door knob, twisting a jar cap, or walking to the bathroom from the bedroom
- Degree of return to the patient's previous sports level

Also, patient-reported experience measures (PREMs) are instruments for real-time feedback about the experience of patients' quality of care. Three domains have been described: patient safety (includes hygiene and physical safety), clinical effectiveness (results of procedures and aftercare), and patient experience (compassion, dignity, respect, etc.). It has been shown there is a weak association between PREM and PROM results: In an English study, this was shown after hip and knee surgery using institutional PREMs and well-validated PROMs (Oxford hip and knee scores) [5].

Patient-reported outcome instruments (PROMs) are important because they show how satisfied patients are with what we do. Sometimes these outcomes are referred to as "subjective" but should generally be avoided due to the misinterpretation of these as "not objective" [14].

However, why are PROMS so important for clinical outcome measurements?

Firstly, it is their ability to measure not only the "objective" goals of a given procedure or treatment but the real impact on the patient's quality of life, which is the most important reason of every treatment.

Secondly, health policy focuses more frequently on the patients' well-being. PROMs are a valuable method to obtaining information directly from the ultimate target of health interventions—the patients—and thus not only measuring the impact of these interventions but guiding us toward new solutions [30].

As stated earlier, patient-reported outcome measurements can also be classified into two groups:

- General health status: They report the patient's overall well-being and functionality; they can be applied to multiple medical etiologies and across multiple patients with different cultural and educational backgrounds [14]
- Disease/joint/region specific: They measure a condition's effect on the patient. Disease-specific measurements focus on a subgroup of patients affected by a condition and can measure the effect of changes in it. They have the advantage of measuring general effects of a disease other than those related to its location. Region specific (or joint specific) measures the effect of any condition that affects a specific part of the body; these measurements can be applied to diverse etiologies that affect a specific region. The disadvantage sometimes is that they cannot differentiate the etiology if more than one is present (e.g., the ASES self-evaluation form score can be affected if the patient has both a rotator cuff injury and a secondary frozen shoulder). They have the advantage that they can measure small changes on a specific region and are relatively easy to apply [14]

Fact Box 10.1

Types of PROMs (subjective own measurement of patient status).

– General health status

For example, Medical Outcomes Study Short Form (SF) 36 [31]

– Disease specific

For example, Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) [4] (hip/knee osteoarthritis)

– Joint/region specific

- Upper limb DASH (disabilities of the arm, shoulder, and hand) index [3]
- Shoulder ASES (American Shoulder and Elbow Society) self-evaluation form score [26]
Shoulder activity level [6]
- Elbow Oxford elbow score [8]
- Hand/wrist: PRWHE (patient-rated wrist/hand evaluation questionnaire) [19]
- Hip Hip disability and osteoarthritis outcome score (HOOS) [24]
Hip outcome score (HOS) [20]
- Knee WOMAC [4]
Knee injury and Osteoarthritis Outcome Score (KOOS) [17, 27]
IKDC (subjective form) [12]
Lysholm score [18]
Tegner activity scale [29]
- Foot and ankle FAAM Foot and Ankle Ability Measure (FAAM) [21]
Foot Function Index (FFI) [7]
Ankle Osteoarthritis Scale (AOS) [9]
- Lumbar Oswestry Low Back Pain Disability Questionnaire [10]

First, the researchers should determine what is evaluated for his/her research purpose specifically. Outcome measures can be selected based on which region, such as the knee, shoulder, or others, and which parameters, such as pain, subjective function, and/or satisfaction.

The first idea which outcome instruments should be used can be taken from previous papers within a similar research topic.

Performance of each outcome measure can be evaluated by psychometric properties, including its reliability, validity, and responsiveness. Reliability is commonly appraised by the consistency and accuracy of the evaluation. Repeatability of the measurement can be estimated by the intraclass correlation coefficient (ICC) for continuous measures or the Cohen coefficient for discrete measures, and greater than 0.8 of those coefficients are generally acceptable. The standard error of a measurement (SEM) is used to estimate the precision of the evaluation. The magnitude of the ICC defines a measure's ability to discriminate among subjects, and the SEM quantifies error in the same units as the original measurement.

Validity refers to how exactly the outcome measure evaluates what it is expected to assess. Each outcome measure has a specific area of evaluation, and the researcher should consider how much the purpose of his/her research would be addressed by the selected outcome measure. For example, KOOS score has a great validity to evaluate the subjective symptom and function of the knee related to the osteoarthritis [17], whereas it demonstrates poor validity for sport-related knee injury and treatment.

Researchers should also consider the responsiveness of the outcome measure. Responsiveness is determined by how much the outcome measure can detect clinically important change over time and/or intervention. If the selected outcome measure has only poor responsiveness, the clinically significant effect would be missed. Minimal detectable change (MDC) and minimally clinically important difference (MCID) are generally used to indicate the responsiveness of an outcome measure.

10.3 How to Select Appropriate Patients' Outcome Instruments for My Study?

There are a variety of outcome instruments, which are frequently used in orthopedic research. When a research study is designed, it is an important step to select the best and most appropriate outcome instrument. When the wrong choice of outcome instruments is used, the scientific value is reduced or at worst diminished.

The MDC is calculated from the ICC and the standard deviation of the outcome measure and implies the range of possible measurement error. If the change of an outcome measure is lower than its MDC, this change can technically be regarded as nothing. Even if the change of an outcome measure is more than MDC, it is still undetermined if the patients perceive improvement or deterioration. The difference over the MCID would be highly likely to affect the patient impression significantly. Each outcome measure has its MDC and MCID [1, 2, 11, 13, 15–17, 22, 23, 25, 28], but they are not always referable in the literature especially for some new outcome measures. The researcher should be aware of the responsiveness of their selected outcome measure and would be better to select those whose MDC and MCID are available.

Further selection process goes through consideration of their practical administration. There are some factors which should be considered, such as usability, applicability, and affordability. Questionnaires should be easy to use and understandable. Only then lay persons such as patients can answer these without any trouble. Language problems might happen in some non-English speaking countries. Because majority of the outcome instrument questionnaires are written in English, proper translation and cross-cultural adaptation of the questionnaires are necessary. A properly translated version of the questionnaire would be very helpful, but they are not always available for some languages. Mistranslation or misunderstanding by the patients might impair the quality of the outcome instrument. If there is no available translated version of the questionnaire, the researcher should make his/her own translation and go through validation process. Also, the scoring should be simple to interpret the results with minimal effort.

When you design a clinical study, at least one PROM and one clinician-based outcome instrument which are widely accepted in your field of interest should be utilized for each study. Otherwise, the results would not be compared to those reported in the previous papers. Rarely used outcome measures might need other addi-

tional outcome measures to assess the clinical and/or scientific impact of the study results.

Most clinical outcome measures do not require administration cost, but, in rare cases, e.g., SF-36, additional cost is needed for collecting and analyzing the data and limits the application of the outcome measure.

Fact Box 10.2

Here are the issues which should be considered to find the best selection of outcome instruments for your study. Technically solid and practically usable PROM can be selected through these items.

- Technical issues
 - Is it reliable?
ICC and SEM should be checked for continuous data while Cohen coefficient for discrete data.
 - Is it valid for your research purpose?
 - Is it responsive enough to answer your research question?
MDC and MCID of the outcome measure should be checked.
- Practical issues
 - Is it commonly used in the same field of research?
 - Is it simple to use and analyze?
 - Is it readable and understandable for your patients?
 - Is it available and affordable for you?

Clinical Vignette

A physician innovated a new postoperative rehabilitation protocol after the ACL reconstruction. He wanted to demonstrate the clinical advantage of the newly developed rehabilitation method and considered a clinical research. The purpose of this research was determined to compare the short-term, i.e., 2-year, clinical outcome after the ACL reconstruction between conventional and

newly developed postoperative rehabilitations in the young athletes. In addition to the clinical evaluations of the knee stability including KT-measurement, IKDC score was collected as a physician-reported outcome. Since the patient population was young and the follow-up was relatively short, Lysholm and IKDC subjective scores were selected for this study as a patient-reported outcome. In this case, KOOS score was supplementary assessed but not reported due to a lack of relevance to this study purpose.

Take-Home Message

- Patient's outcome measures are requisite for every clinical research.
- There are a variety of choices for the patient's outcome measures, and the researchers should select just the right set of them to address their study purpose.
- Appropriate selection of the outcome measure goes through clarifying the study targets, listing up practically usable ones, and deciding one or more depending on their reliability, validity, and responsiveness.

References

1. Ansari NN, Naghdi S, Hasanvand S, Fakhari Z, Kordi R, Nilsson-Helander K. Cross-cultural adaptation and validation of Persian Achilles tendon Total Rupture Score. *Knee Surg Sports Traumatol Arthrosc.* 2016;24(4):1372–80.
2. Barber-Westin SD, Noyes FR, McCloskey JW. Rigorous statistical reliability, validity, and responsiveness testing of the Cincinnati Knee Rating System in 350 subjects with uninjured, injured, or anterior cruciate ligament-reconstructed knees. *Am J Sports Med.* 1999;27:402–16.
3. Beaton DE, Katz JN, Fossel AH, Wright JG, Tarasuk V, Bombardier C. Measuring the whole or the parts? Validity, reliability, and responsiveness of the Disabilities of the Arm, Shoulder and Hand outcome measure in different regions of the upper extremity. *J Hand Ther.* 2001;14:128–46.
4. Bellamy N, Watson Buchanan W, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: A health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol.* 1988;15:1833–40.
5. Black N, Varaganam M, Hutchings A. Relationship between patient reported experience (PREMs) and patient reported outcomes (PROMs) in elective surgery. *BMJ Qual Saf.* 2014;23:534–42.
6. Brophy RH, Beauvais RL, Jones EC, Cordasco FA, Marx RG. Measurement of shoulder activity level. *Clin Orthop Relat Res.* 2005;439:101–8.
7. Budiman-Mak E, Conrad KJ, Roach KE. The Foot Function Index: a measure of foot pain and disability. *J Clin Epidemiol.* 1991;44:561–70.
8. Dawson J, Doll H, Boller I, Fitzpatrick R, Little C, Rees J, et al. The development and validation of a patient-reported questionnaire to assess outcomes of elbow surgery. *J Bone Joint Surg Br.* 2008;90(4):466–73.
9. Domsic RT, Saltzman CL. Ankle Osteoarthritis Scale. *Foot Ankle Int.* 1998;19:466–71.
10. Fairbank JCT, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy.* 1980;66:271–3.
11. Greco NJ, Anderson AF, Mann BJ, Cole BJ, Farr J, Nissen CW, et al. Responsiveness of the International Knee Documentation Committee Subjective Knee Form in comparison to the Western Ontario and McMaster Universities Osteoarthritis Index, modified Cincinnati Knee Rating System, and Short Form 36 in patients with focal articular cartilage defects. *Am J Sports Med.* 2010;38(5):891–902.
12. Irrgang JJ, Anderson AF, Boland AL, Harner CD, Kurosaka M, Neyret P, et al. Development and validation of the International Knee Documentation Committee Subjective Knee Form. *Am J Sports Med.* 2001;29:600–13.
13. Irrgang JJ, Anderson AF, Boland AL, Harner CD, Neyret P, Richmond JC, et al. International Knee Documentation Committee. Responsiveness of the International Knee Documentation Committee Subjective Knee Form. *Am J Sports Med.* 2006;34(10):1567–73.
14. Irrgang JJ, Lubowitz JH. Measuring arthroscopic outcome. *Arthroscopy.* 2008;24(6):718–22.
15. Kemp JL, Collins NJ, Roos EM, Crossley KM. Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. *Am J Sports Med.* 2013;41(9):2065–73.
16. Kocher MS, Steadman JR, Briggs KK, Sterett WI, Hawkins RJ. Reliability, validity, and responsiveness of the Lysholm knee scale for various chondral disorders of the knee. *J Bone Joint Surg Am.* 2004;86-A(6):1139–45.
17. Lyman S, Lee YY, Franklin PD, Li W, Cross MB, Padgett DE. Validation of the KOOS, JR: a short-form knee arthroplasty outcomes survey. *Clin Orthop Relat Res.* 2016;474(6):1461–71.
18. Lysholm J, Gillquist J. Evaluation of knee ligament surgery results with special emphasis on use of a scoring scale. *Am J Sports Med.* 1982;10:150–4.

19. MacDermid JC, Turgeon T, Richards RS, Beadle M, Roth JH. Patient rating of wrist pain and disability: a reliable and valid measurement tool. *J Orthop Trauma*. 1998;12(8):577–86.
20. Martin RL, Philippon MJ. Evidence of validity for the Hip Outcome Score in hip arthroscopy. *Arthroscopy*. 2007;23:822–6.
21. Martin RL, Irrgang JJ, Burdett RG, Conti SF, Van Swearingen JM. Evidence of validity for the Foot and Ankle Ability Measure (FAAM). *Foot Ankle Int*. 2005;26:968–83.
22. Marx RG, Jones EC, Allen AA, Altchek DW, O'Brien SJ, Rodeo SA, et al. Reliability, validity, and responsiveness of four knee outcome scales for athletic patients. *J Bone Joint Surg Am*. 2001;83-A(10):1459–69.
23. Mohtadi NG, Griffin DR, Pedersen ME, Chan D, Safran MR, Parsons N, et al. Multicenter Arthroscopy of the Hip Outcomes Research Network. The Development and validation of a self-administered quality-of-life outcome measure for young, active patients with symptomatic hip disease: the International Hip Outcome Tool (iHOT-33). *Arthroscopy*. 2012;28(5):595–605.
24. Nilsson AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)—validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord*. 2003;4:10.
25. Papou A, Hussain S, McWilliams D, Zhang W, Doherty M. Responsiveness of SF-36 Health Survey and Patient Generated Index in people with chronic knee pain commenced on oral analgesia: analysis of data from a randomised controlled clinical trial. *Qual Life Res*. 2017;26(3):761–6.
26. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, et al. A standardized method for the assessment of shoulder function. *J Shoulder Elb Surg*. 1994;3:347–52.
27. Roos EM, Lohmander LS. The Knee injury and Osteoarthritis Outcome Score (KOOS): from joint injury to osteoarthritis. *Health Qual Life Outcomes*. 2003;1:64.
28. SooHoo NF, Li Z, Chenok KE, Bozic KJ. Responsiveness of patient reported outcome measures in total joint arthroplasty patients. *J Arthroplast*. 2015;30(2):176–91.
29. Tegner Y, Lysholm J. Rating systems in the evaluation of knee ligament injuries. *Clin Orthop Relat Res*. 1985;198:43–9.
30. Walton MK, Powers JH III, Hobart J, Patrick D, Marquis P, Vamvakas S, et al. International Society for Pharmacoeconomics and Outcomes Research Task Force for Clinical Outcomes Assessment. Clinical Outcome Assessments: Conceptual Foundation—Report of the ISPOR Clinical Outcomes Assessment—Emerging Good Practices for Outcomes Research Task Force. *Value Health*. 2015;18(6):741–52.
31. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30:473–83.



Basics of Outcome Assessment in Clinical Research

11

Monique C. Chambers, Sarah M. Tepe,
Lorraine A. T. Boakye, and MaCalus V. Hogan

11.1 Introduction

The [World Health Organization](#) defines an outcome measure as the “change in the health of an individual, group of people, or population that is attributable to an intervention or series of interventions” [13]. Measuring clinical outcomes is essential to track progress as healthcare institutions continue to shift toward more value-based policies, programs, and practice. Outcome measures are largely motivated by national standards of care and financial incentives to minimize waste and cost burden associated with care delivery. Clinical outcome measures may include health-related factors, health system-related factors that serve as a proxy for medical resource utilization, and/or the subjective perspective of patients. Tracking clinical outcomes allows healthcare organizations to identify variations in care delivery to properly align healthcare practices within an organization. Additionally, it provides a model to assess evidence-based practices and determine which

medical interventions are best suited for the local patient population. Finally, transparency in clinical outcomes empowers healthcare institutions to make healthcare decisions that help to improve patient care and ultimately drive down cost.

There has been an increased focus on various outcome measures because of clinical interventions. Much of the focus on clinical and translation research seeks to satisfy the scientific curiosity of the clinician and/or scientist. However, as research guides clinical management, it is important to apply the appropriate scientific methods to accurately assess how medical interventions impact patient outcomes. In this chapter, we will discuss these outcome measures to guide clinicians that are pursuing research and to equip young investigators with the necessary tools to adequately design studies that utilize the appropriate measure to assess outcomes based on the clinical problem.

11.2 Principles of Outcome Measures

An outcome measure is the result of a test that is used objectively to determine the baseline function of a patient at the beginning of treatment [13]. Once treatment has commenced, the same metric can be used to determine progress and

M. C. Chambers
Foot and Ankle Injury Research [F.A.I.R.] Group,
Department of Orthopaedic Surgery, University of
Pittsburgh School of Medicine, Pittsburgh, PA, USA

S. M. Tepe · L. A. T. Boakye · M. V. Hogan (✉)
Department of Orthopaedic Surgery, University of
Pittsburgh School of Medicine, Pittsburgh, PA, USA
e-mail: hoganmv@upmc.edu

treatment efficacy. Appropriate outcome measures allow one to assess the quality of medical interventions. The most useful outcomes should be measurable based on discrete parameters and distinct time points. The measurement tool should be accurate within a small range of error. The outcome measure should also be reliable with the ability to achieve the same result if applied again by a well-trained provider. The tool should be validated to ensure consistency with other measures used to track outcomes [10]. As such, the goal of measuring outcomes is to provide a baseline assessment for the improvement of patient outcomes, patient satisfaction, and the cost burden associated with healthcare.

11.3 Considering Various Data Sources

Regulatory agencies seek to streamline the collection and utilization of data sources; therefore,

clinical investigators should understand which data sources are best suited to answer their clinical question. Mandates related to the meaningful use of electronic medical records lead the Food and Drug Administration (FDA) to establish aims of electronic data sources. It is also important to understand how various data sources are generated. Common data sources are generated from administrative claims and health insurance enrollment, clinical encounters with providers, patient surveys, and clinical trials, among other sources. There are many advantages/disadvantages to these sources (Table 11.1). The appropriate collection, maintenance, and utilization of data sources will result in the accurate assessment of medical treatments and outcomes that are specific to the patient population of interest. Clinical investigators should have a critical eye for the data source used and should seek to constantly improve the thorough collection of medical data and outcomes assessment.

Table 11.1 Understanding data sources [14]

Data sources	Advantages	Disadvantages
Administrative data	<ul style="list-style-type: none"> Standardized coding systems Electronic accessibility Easy to access when de-identified 	<ul style="list-style-type: none"> Intent of data is for billing, rather than quality reporting/care Concerns with accuracy
Medical records	<ul style="list-style-type: none"> Most comprehensive view of medical picture Electronically available across most practices 	<ul style="list-style-type: none"> Costly to obtain and maintain data protection Risk of data breach
Patient surveys	<ul style="list-style-type: none"> Focuses on patient experience as a measure for outcomes/satisfaction Questionnaires, such as PROs, that have been validated to reflect quality outcomes Results may be better understood by patients AND providers 	<ul style="list-style-type: none"> Patients may misunderstand the questions Responses may differ based on the time within the episode of care Need for questionnaire implementation and administration which may be costly Sampling bias of activated patients
Standardized clinical data	<ul style="list-style-type: none"> More comprehensive data set Data set already exist Assess outcomes across multiple domains of cares 	<ul style="list-style-type: none"> May have incomplete data May reflect different medical practices based on regional population standards
Shared data/national databases [16]	<ul style="list-style-type: none"> May enhance understanding of results of individual trials Allows pooling of multiple studies to expand the scientific rigor and implications Improves medical accuracy 	<ul style="list-style-type: none"> May lead to inaccurate conclusions Lack of homogeneity in the populations across studies Endpoints may be defined differently across studies

Fact Box 11.1: The FDA Aims for Electronic Data Sources [15]

- Eliminate unnecessary duplication of data
- Reduce the possibility for transcription errors
- Encourage entering source data during a subject's visit
- Facilitate remote monitoring of data
- Promote real-time access for data review
- Facilitate the collection of accurate and complete data

11.4 Qualitative Methods

A primary area of interest with data sources deals with the use of medical data to assess qualitative outcomes. An increase in qualitative studies came about following the development of the Patient-Centered Outcomes Research Institute (PCORI) in 2010. The PCORI was funded by the Congress to help patients make informed decisions regarding their healthcare decisions. Qualitative research methods prioritize patient needs, experiences, and perspectives. Qualitative research provides clarity to clinical practice in ways that are seldom revealed through quantitative methodologies [12]. Qualitative studies may be conducted through one-on-one interviews, focus group studies, or surveys.

Fact Box 11.2: Qualitative Research

- Seeks to ask a question
- Emphasizes people and process
- Acknowledges that how clinical practice occurs is more important than doing a task
- Results should be placed within the context of the qualitative research question

11.4.1 Clinical Quality Measures (CQMs)

Qualitative research also seeks to address inadequacies in quality of care that cannot be directly explained by large data pools. A focus on quality improvement projects promotes accountability and transparency in healthcare outcomes. Clinical quality outcome measures help establish the standard of care by measuring or quantifying healthcare processes, outcomes, and organizational structure. Healthcare organizations and clinical researchers use parameters such as mortality, readmission, and complications as a measure to improve both high-quality care and cost goals. These goals include care that is effective, safe, efficient, patient-centered, equitable, and timely care [6]. Additionally, the Centers for Medicare and Medicaid Services (CMS) uses several clinical quality measures to calculate overall hospital quality performance.

Fact Box 11.3: Clinical Quality Measures (CQMs) of Hospital Performance [6]

- Mortality (22%)
- Safety of care (22%)
- Readmissions (22%)
- Patient experience (22%)
- Effectiveness of care (4%)
- Timeliness of care (4%)
- Efficient use of medical imaging (4%)
- Other MSK complications

11.5 Performance-Based Outcome Measures

Performance-based outcome measures have been used for decades in orthopedics including in biomechanical and kinematic studies. Performance-based outcome measures usually involve observation of the patient in some capacity to assess functional status. The advantage of collecting objective data means that the results should be reproducible and without bias [16].

Despite the widespread use of performance based outcomes, many have not been validated or correlated with clinical outcomes. Clinical outcome measures may be applied inappropriately to a population that the outcome measures were not designed or validated to assess. Such measures should be modified and validated for the current population of interest prior to use in a clinical investigation. Nonetheless, performance-based outcomes present an objective method to determine a patient's functional capacity.

11.6 Self-Reported Outcome Measures

Outcomes in healthcare are not only representative of the quality of care provided but also reflect the service as experienced by the patient [1]. Self-reported outcome measures were developed as a means of gathering information based on patient experience and perception. Many of these measures was designed to assess treatment effectiveness and to emphasize the importance of the individual's own perspective when making the evaluation of care [2, 8]. These outcomes may include pain scores, patient satisfaction scores, as well as patient-reported outcomes (PROs). The challenge is to encourage healthcare professionals (HCPs) to focus on individual patient perceptions of disease impact as well as disease activity measures, with the aims of providing patient-centered care through shared decision-making and improving outcomes considered important by patients themselves [3].

11.7 Patient-Reported Outcome Measures

11.7.1 Development of PROs

PROs are often a hybrid of performance-based and self-reported measures that provide a more holistic picture of the patient's activity and satisfaction. PROs are standardized, validated questionnaires that are completed by patients to measure their perceptions of their own functional

status and well-being [2]. The collection of PROs has become more prevalent in physician practices across the United States. The use of PROs is a valuable tool for physicians, patients, and clinical investigators. PROs have been collected by physicians for many years; however, they have historically been used only to determine the efficacy of various treatments [2].

In the past few years, the focus of a patient's self-assessment has shifted towards the idea that the patient perspective and input is an integral aspect of clinical assessment and progress. Patients can actively participate in medical decision-making to help guide their treatment and understand improvement or changes that may occur. Each patient responds to treatment in a unique way. Therefore, it is important that, while standardized, the questionnaires are answered accurately by the patient. While physicians remain the educators, patients can maintain a certain degree of autonomy by self-reporting symptoms, progress, and health-related quality of life [3]. Medical professionals, researchers, hospital administrators, and even insurers still use the information as a guide to what treatments are most clinically and financially effective. The constant progression toward value-based policies and implementation will lead to increased value of PRO utilization, Orthopaedic surgery will likely be impacted especially given the higher rate of elective procedures within our field [4].

11.7.2 Collecting PROs Within a Systematic Performance Platform

The collection of PROs has traditionally been managed via telephone calls or clinical visits. Data were recorded in paper charts and then later manually abstracted or transcribed into a clinical database. This system was an inefficient way for systems to track the completeness and accuracy of outcomes collection. The meaningful use mandate within the Patient Protection and Affordable Care Act (ACA) incentivized health-care systems to not only implement an appropriate electronic medical record (EMR) but to also

establish systems designed to optimize data collection, analysis, and interpretation. The use of the EMR affords the opportunity to optimize PRO data collection.

The development of a system-wide performance platform allows for uniformity in the collection process. To optimize PRO collection, there should be a standardized system in place for patients to complete specific PRO questionnaires at distinct time periods within the disease process and an episode of care. The platform can be built into the EMR database with predetermined or “red flag” reminders to alert healthcare providers when patients are due for the next set of surveys. This information is then stored within the EMR and can be pulled as part of a large data set to assess the impact on the patient population. This information can also be retrieved and viewed as part of the individual patient’s EMR to track the patient’s progress over time. Access to the PRO results over time allows the physician to use the patient’s daily activities and overall health status in the shared decision-making process for treatment.

Despite the vast benefits of collecting and utilizing PROs within the clinical and research settings, there remain many challenges that prevent PROs from being implemented across the

healthcare system. In order to ensure the accurate assessment of PRO data, the information should be thorough, complete, and collected at key stages within the disease state [7]. Any variations within the data set create limitations to both the use and the validity of outcomes data.

11.7.3 Utilization of PROs

Patient-reported outcomes create an environment that fosters shared decision-making. Decisions made with patient input often make that patient feel more at ease because he or she can take on a more active role in regards to guiding treatment. Although tracking PROs has become more commonplace in orthopaedic research, there has been difficulty in translating the results into changes within clinical practice. To best serve as a guide for clinical decision-making, the collection and assessment of PRO outcomes should take clinical relevance into account. To accomplish this goal, specific PROs should be validated based on the specific orthopedic disorder or disease process, particularly for PROs that measure general health status, which may or may not be greatly impacted by the orthopedic condition.

While this example is not the direct result of a research study, the continuous use of PROs in the clinical decision-making process can help to shape clinical guidelines and impact the results of larger population studies as it relates to the utilization of PROs. This type of clinical relevance should be within the aims of clinical investigations related to all clinical outcome measures. With clinical implications at the forefront of outcome research study design, researchers are better equipped to develop the right questions and collect the appropriate data to address the root cause of outcome discrepancies.

11.7.4 Linking PROs to Comparative Effectiveness Research

To improve the efficiency of PRO collection, the EMR provides new opportunities to have meaningful use of the data collected and pooled in ways that

Clinical Vignette

A patient with an acute ankle injury may have SF-12 and PROMIS-10 scores that reflect poor health status at baseline. If this patient were to undergo surgical intervention, there may be a limited change in their SF-12 or PROMIS-10 scores. This patient would be more accurately assessed based on foot- and ankle-specific outcome measures, such as the AOFAS, GRC, and FAAM scores. Additionally, if this patient were to be included in a study that assessed scores at 3-, 6-, and 12-month follow-up, there would be a large selection bias. Therefore, the *change* in PRO measures as the result of a medical intervention is equally, if not more, important to evaluate as the raw pre- and post-operative scores.

are innovative and necessary for the shared decision-making process. Comparative effectiveness research (CER) attempts to synthesize current medical investigations on various medical interventions with the goal of identifying the optimal management plan for several orthopedic injuries [11]. As such, the purpose of CER is multifaceted and aligns well with the utilization of PRO data. Linking PRO results provides an evidence-based approach to the diagnosis, treatment, prevention, and monitoring of orthopaedic conditions [5]. A comprehensive approach must be adopted to ensure that the diverse patient population has been accounted for in medical research. PROs that account for diversity within our population is the only way to effectively educate patients on which treatment has the most appropriate outcomes for their specific condition and demographics. Unfortunately, many patient populations that would be impacted by such medical interventions are excluded from or opt out of participation in scientific studies. As a scientific investigator and clinical

practitioner, there should be consideration for the clinical implications that can be applied based on the results of any given study. A better understanding of the added value of PROs for CER will help to develop best practices and define more formally when PROs have demonstrated utility [9].

Take-Home Message

- The shift from volume-based, physician-focused care to value-based, patient-centered models has sparked a change in the way treatment effectiveness is assessed.
- The use of clinical outcome measures that reflect the true value of high-quality and cost-efficient care should be considered when evaluating medical interventions and determining if treatment modifications are necessary in distinct patient populations.
- These principles should serve as a guide to the basis of outcome assessment and study design for clinical outcome investigations.

Clinical Vignette

A 54-year-old patient who is moderately active and has developed ankle arthritis following an ankle arthroscopy for a prior osteochondral lesion of the talus may be considering an ankle fusion. With the appropriate tracking of regionally specific PRO scores, such as the Foot and Ankle Ability Measure (FAAM), American Orthopedic Foot and Ankle Society Score (AOFAS), and Global Rate of Change, results may reveal a drastic improvement in the patients' functional status as it relates to ankle performance. Additionally, general health-related outcome scores, such as short function-12 or PROMIS-10, may also suggest that the patient is functionally improved and perhaps within a range that surgical intervention may have little to no improvement on his/her functional status or well-being. For this patient, an ankle fusion or arthroplasty may be the wrong choice and lead to greater morbidity and higher medical resource utilization without adequate clinical benefit.

Fact Box 11.4: Patient-Reported Outcomes

- PRO data should be thorough, complete, and collected at key stages within the disease state
- Linking PRO results provides an evidence-based approach to the diagnosis, treatment, prevention, and monitoring of orthopedic conditions
- Better understanding of the added value of PRO for CER will help to develop best practices

11.8 Useful Resources

- Kane, RL. Jones & Bartlett Learning's Understanding Health Care Outcomes Research, 2nd Edition. 2006.
- <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/index.html>.
- Keith D. Baldwin; Rachel L. Slotcavage; G. Russell Huffman. Orthoepidemiology 101: The Basics of the Art of Outcomes Research in Orthopaedic Surgery. UPOJ. Vol 19.

References

1. AAOS. Principles of patient reported outcome measures (PROMs) reporting: information statement. 2015. [https://aaos.org/uploadedFiles/PreProduction/About/Opinion_Statements/advistmt/1044%20Principles%20of%20Patient%20Reported%20Outcome%20Measures%20\(PROMs\)%20Reporting.pdf](https://aaos.org/uploadedFiles/PreProduction/About/Opinion_Statements/advistmt/1044%20Principles%20of%20Patient%20Reported%20Outcome%20Measures%20(PROMs)%20Reporting.pdf). Accessed 19 Nov 2017.
2. Dawson J, Doll H, Fitzpatrick R, Jenkinson C, Carr AJ. The routine use of patient reported outcome measures in healthcare settings. *BMJ*. 2010;340:c186. <https://doi.org/10.1136/bmj.c186>.
3. Fautrel B, Alten R, Kirkham B, de la Torre I, Durand F, Barry J, et al. Call for action: how to improve use of patient-reported outcomes to guide clinical decision making in rheumatoid arthritis. *Rheumatol Int*;2018. <https://doi.org/10.1007/s00296-018-4005-5>.
4. Ferguson C, Rocha JL, Lalli T, Irrgang JJ, Hurwitz S, Hogan MV. Developing performance and assessment platforms in foot and ankle surgery. *Foot Ankle Int*. 2016;37(6):670–9.
5. Health Services Research Information Central. Comparative effectiveness research. <https://www.nlm.nih.gov/hsrinfo/cer.html>. Accessed 21 Mar 2018.
6. <https://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/QualityMeasures/index.html>. Accessed 18 Feb 2018.
7. MOTION Group. Current concepts review: patient-reported outcomes in orthopaedics. *J Bone Joint Surg Am*. 2018;100:436–42.
8. Rolfson O, Rothwell A, Sedrakyan A, Chenok KE, Bohm E, Bozic KJ, et al. Use of patient-reported outcomes in the context of different levels of data. *J Bone Joint Surg Am*. 2011;93(Suppl 3):66–71. <https://doi.org/10.2106/JBJS.K.01021>.
9. Segal C, Holve E, Sabharwal R. Collecting and using patient-reported outcomes (PRO) for comparative effectiveness research (CER) and patient-centered outcomes research (PCOR): challenges and opportunities. 2013. <http://www.academyhealth.org/files/Collecting%20and%20Using%20PRO%20for%20CER%20and%20PCOR.pdf>.
10. Selecting the Right Measures for Your Report. Agency for healthcare research and quality. <https://www.ahrq.gov/professionals/quality-patient-safety/talkingquality/create/gather/index.html>. Accessed 16 Mar 2018.
11. Slutsky J, Atkins D, Chang S, et al. Comparing medical interventions: AHRQ and the effective health care program. In: *Methods guide for effectiveness and comparative effectiveness reviews* [Internet]. Rockville: Agency for Healthcare Research and Quality; 2008.
12. Smythe L, Giddings LS. From experience to definition: addressing the question: “what is qualitative research?”. *Nurs Prax N Z*. 2007;23:37–57.
13. Tinker A. The top 7 outcome measures and 3 measurement essentials. *HealthCatalyst*. <https://www.healthcatalyst.com/The-Top-7-Outcome-Measures-and-3-Measurement-Essentials>. Accessed 17 Feb 2008.
14. Understanding Data Sources. Agency for healthcare research and quality. <https://www.ahrq.gov/professionals/quality-patientsafety/talkingquality/create/understand.html>. Accessed 21 March 2018.
15. US Food and Drug Administration. Guidance for industry: electronic source data in clinical investigations. <http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm328691.pdf>. Accessed 17 Feb 2018.
16. Wang S, Hsu CJ, Trent L, Ryan T, Kearns NT, Civillico EF, et al. Evaluation of performance-based outcome measures for the upper limb: a systematic review. *PM&R*. 2018;10(9):951–962.e3.



Types of Scoring Instruments Available

12

José F. Vega and Kurt P. Spindler

12.1 Introduction

Research, of any kind, requires the measurement of an outcome. Possibly the earliest recorded account of “research” dates back to 550 BC. According to the “Book of Daniel,” of *The Bible*’s Old Testament, the Babylonian King Nebuchadnezzar mandated that all Babylonians consumed a diet exclusively of meat and wine, as this would keep them in excellent health. A handful of young men of royal lineage with herbivorous inclinations objected to this diet. The king decreed that these young men would eat only beans and water for 10 days, after which he would assess their level of nourishment. To the king’s surprise, he found the legume lovers to be better nourished than those sustained on meat and wine, and, thus, he allowed the diet to continue [7, 14, 69, 81].

Research and outcome measurement has certainly come a long way from legumes and nourishment, but rigorous clinical research is still in its early years, relatively speaking. More than

two millennia would pass between King Nebuchadnezzar’s legume experiment and the earliest examples of clinical research. It was not until the turn of the twentieth century that Dr. Ernest Amory Codman, considered by many to be the father of outcomes research, began advocating his “end result” system, which he famously described as “The common sense notion that every hospital should follow every patient it treats, long enough to determine whether or not the treatment has been successful, and then to inquire, ‘If not, why not?’ with a view to preventing similar failures in the future [9].” This idea (which was received quite poorly by his peers and ultimately cost him his job at the Massachusetts General Hospital) led Codman to develop the first registry, which he prophetically envisioned as a means to implement his “end result” idea as standard of care on a national level [13].

It would take another 40 years from Codman’s development of the “end result” to the first randomized controlled trial, in which streptomycin was compared with placebo in the treatment of pulmonary tuberculosis, which took place in 1946 [7]. Nearly another 50 years would pass before Gordon Guyatt, a young internist from McMaster University in Ontario, Canada, coined the term “evidence-based medicine,” and modern orthopedic research, which has come to rely more and more heavily on PROMs, is younger still [16, 81].

J. F. Vega
Cleveland Clinic Lerner College of Medicine,
Cleveland, OH, USA
e-mail: vegaj@ccf.org

K. P. Spindler (✉)
Cleveland Clinic Sports Health Center,
Garfield Heights, OH, USA
e-mail: spindlk@ccf.org, stojsab@ccf.org

Fact Box 12.1

The first mention of the phrase “evidence-based medicine” in a major American medical journal occurred in the November 1992 edition of the *Journal of the American Medical Association* (JAMA). In 2001, a mere 9 years after its debut, the phrase appeared in >2500 publications.

outcome measure alone would not have provided such a complete picture demonstrating no difference between the BTB and ST/G grafts in this population. The importance of utilizing multiple outcome measures of both varieties is underscored by the use of both types of data to perform large-scale systematic reviews and meta-analyses [62, 68].

12.2 Types of Scoring Instruments: Patient-Reported or Observer-Recorded?

Two major categories of scoring instruments exist for use in clinical research: PROMs and observer-recorded outcome measures (OROMs). The former rely exclusively on patient input (e.g., chronicity, frequency, severity, and impact of symptoms, instability, limitations, etc.), while the latter take into account that which an observer can measure (e.g., knee laxity, degree of articular cartilage damage, functional ability, etc.). Neither is superior or inferior to the other, but instead, the two work hand in hand to provide a complete picture of a patient’s status.

Clinical Vignette 1

To illustrate the utility of PROMs in conjunction with OROMs, consider a 2006 study conducted by Svensson et al. The authors compared PROMs, instrumented knee laxity, and results of a functional testing in female patients who had undergone anterior cruciate ligament reconstruction (ACL-R) using either bone-patellar tendon-bone (BTB) or four-strand semitendinosus/gracilis (ST/G) autografts. At 2-year follow-up, the investigators found no significant differences between the groups with regard to PROMs, instrumented laxity (using the KT-1000 arthrometer), or function (one-leg-hop test) [70]. The use of one

12.3 Patient-Reported Outcome Measures: General, Joint-Specific, and Generic Options

Orthopedics as a whole, and sports medicine in particular, is a field in which traditional outcome measures, such as mortality, are of little utility. This is because orthopedics (and, again, orthopedic sports medicine in particular) radically impacts *quality* of life—more so than *quantity* of life. Unfortunately, quality of life is often much more difficult to measure and is an outcome that hinges on many dimensions (e.g., pain, function, limitations, etc.). This challenge is not unique to orthopedics, and, thus, research into the development and use of PROMs has grown considerably over the last two decades [25].

PROMs are validated questionnaires completed by patients to generate a score that can be tracked over time (to observe a change in a specific patient) and/or compared with scores from other patients. While initially designed for use in clinical research, they are being included more and more frequently as a routine part of patient care, as the results are readily interpreted by health-care providers and can be used to help direct patient care.

PROMs come in a multitude of varieties, including general health measures, joint-specific measures, and more generic measures. The list of available PROMs is long and continues to grow almost every day. Rather than attempt to provide a comprehensive review of all PROMs, the coming pages will review the PROMs most commonly used in the current literature.

Fact Box 12.2

Over the past decade, PROMs have become increasingly more present as not only a research tool but also a metric by which performance can be measured. The use of PROMs is steadily becoming a “standard of care” at multiple institutions across the globe, as PROM data from individuals is used to guide treatment decisions, while aggregate PROM data is used to keep providers accountable and to measure their performance. In 2015, the AAOS Board of Directors convened the Quality Outcomes Data (QOD) Work Group to further investigate and evaluate PROMs.

12.4 General Health PROMs

General health PROMs are designed to quantify a patient’s overall state of health including both physical health and mental health. Three of the most commonly used general health PROMs include the Veterans RAND 12 (VR-12), the PROMIS Global 10, and the EuroQol-5D (EQ-5D).

12.4.1 The VR-12

The VR-12 was developed by the RAND corporation (the name is a shortening of *research and development*) as a by-product of the Medical Outcomes Study, a cross-sectional study involving over 22,000 patients with the goal of developing practical tools to monitor patient outcomes [71]. The Medical Outcomes Survey resulted in the development of the 36-item Short-Form Health Survey (SF-36). The SF-36, which features generic quality-of-life measures, was developed with the goal of using the score to explain variations in patient outcomes (e.g., lower SF-36 scores indicate “less healthy” populations that might do poorly following a certain intervention due to their poor health status rather than the intervention itself). A few small changes were made to the SF-36, and this “new”

questionnaire was then administered to nearly 2500 US military veterans to develop the Veterans RAND-36 (VR-36) [37, 38]. The VR-12 is derived from the VR-36 and, despite being only a third of the length of its parent survey, yields accurate estimates of the VR-36. The VR-12 measures health-related quality of life across seven domains: physical functioning, role limitations due to physical and emotional problems, bodily pain, energy/fatigue, social functioning, mental health, and general health. The results of the VR-12 are reported as two scores—a physical component score (PCS) and a mental component score (MCS) [44].

Another commonly used general health PROM is the SF-12 (a condensed version of the SF-36). The benefit of the VR-12 over the SF-12 is that it is readily available (the rights to the SF-12 are owned by a private corporation currently), utilizes a five-point rating scale based on item response theory (compared with the SF-12, which utilizes “yes/no” options; this helps to mitigate floor and ceiling effects), and also includes two questions to assess patient perception of change in health status over time—something that is very important. Like the VR-12, the SF-36 is also available in the public domain but, again, is lengthier and does not provide substantial more information than the VR-12.

Clinical Vignette 2

The utility and importance of general health PROMs such as the VR-12 are illustrated by their use in large, randomized controlled trials. A recent publication in the *New England Journal of Medicine* utilized the VR-12 to assess the impact of aggressive blood pressure control on quality of life (utilizing data from the Systolic Blood Pressure Intervention Trial [SPRINT]). The authors concluded that aggressive blood pressure control did not negatively impact PROMs as measured by the VR-12 [6].

12.4.2 PROMIS Global 10

The Patient-Reported Outcomes Measurement Information System (PROMIS) Global 10 is another general health PROM and was developed by the NIH in 2004. It is a ten-question survey that measures physical functioning, fatigue, pain, emotional distress, and social health [29]. Bridges have been developed to allow comparisons between the PROMIS 10 Global and VR-12 PCS and MCS [63]. These are algorithms that allow for the conversion of the VR-12 PCS and MCS scores to the PROMIS 10 Global scale so that they can be directly compared.

12.4.3 EuroQol-5D

The EuroQol-5D (EQ-5D) was developed in the late 1980s by an interdisciplinary five-country group that set out to develop a brief, general health measurement tool. They ultimately settled on five questions, each of which addresses a single dimension: mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Each question has five levels [19].

Fact Box 12.3

As of 2009, the use of the EQ-5D is considered a standard of care by England's National Health Service (NHS) and is being administered pre- and postoperatively to all patients undergoing elective total hip and total knee arthroplasty (alongside a joint-specific PROM) [11].

12.5 Joint-Specific Patient-Reported Outcome Measures

While useful and informative, general health PROMs are not used as primary end points in most orthopedic research because, when used in isolation, they lack the responsiveness needed to assess the true impact of an orthopedic intervention.

For example, the VR-12, PROMIS Global 10, and EuroQol-5D have been compared with one another in patients undergoing knee arthroscopy, and all three were found to have equal responsiveness. However, none of these general health outcome measures were as responsive as the Knee injury Osteoarthritis Outcome Score (KOOS), a joint-specific PROM described below [54]. This supports the notion that, in orthopedics, general health outcome measures should always be accompanied by a joint-specific PROM.

What follows is a brief discussion of joint-specific PROMs. As mentioned previously, this list is not meant to be exhaustive but rather features the joint-specific PROMs that can be used to collect quality orthopedic data.

12.5.1 Oswestry Disability Index (ODI; Spine)

The Oswestry Disability Index (ODI) is a ten-question survey developed in 1980 for use in patients with back pain. The ten questions are scored from 0 to 5 and assess intensity of pain, ability to care for oneself, ability to walk, ability to sit, ability to stand, ability to travel, sexual function, social life, and sleep quality [23]. Lower scores indicate less disability. The ODI is considered by many to be the gold standard for assessing the impact of back pain. This notion is supported by the use of the ODI in large randomized controlled trials and in meta-analyses [64, 80].

12.5.2 Neck Disability Index (NDI; Spine)

The Neck Disability Index is a modified version of the ODI that was designed for use in patients with cervical spine complaints and shares many of its characteristics. It has the same length (ten questions) and is scored in the same fashion. Like the ODI, NDI is one of the most widely used PROMs for neck health. It is commonly used in

large randomized trials and as a primary outcome in meta-analysis [31, 58]. Also like the ODI, NDI has been demonstrated to have acceptable psychometric properties [27, 73].

12.5.3 Disabilities of the Arm, Shoulder, and Hand Score (DASH) and Quick-DASH (Hand, Wrist, and Elbow)

The Disabilities of the Arm, Shoulder, and Hand Score (DASH) was developed by the Upper Extremity Collaborative Group (EUCG), a joint initiative made up of members from AAOS, the Council of Musculoskeletal Specialty Societies (COMSS), and the Institute for Work and Health. It debuted in 1996 and was designed to evaluate disorders of the upper limb, from the shoulder to the hand [2, 32]. The DASH consists of 38 questions, each of which is scored on a five-point Likert scale. Lower scores indicate less impairment. What makes the DASH rather unique is that patients are assessed on their ability to complete a task regardless of which limb is needed to perform that action. This is considered both a strength and limitation of the instrument.

The Quick-DASH is an abbreviated version that was developed and released in 2005. The Quick-DASH includes only 11 questions but still maintains excellent correlation with its parent survey [4].

12.5.4 American Shoulder and Elbow Surgeons Standardized Shoulder Outcome Score (ASES; Shoulder)

The American Shoulder and Elbow Surgeons Standardized Shoulder Outcome Score (ASES) is the second half of a dual outcome measure. The first half of the assessment includes a set of physician-rated questions. These are typically not reported in the literature. The second half, which is completed by patients, consists of ten

questions scored from 0 to 3 and assesses pain, instability, and ability to perform activities of daily living with the affected shoulder. The total score is converted to a 100-point scale, with higher scores indicating better outcomes [59]. The ASES shoulder outcome score has been validated for use in patients with a wide range of shoulder diagnoses, including shoulder instability [50].

12.5.5 Western Ontario Shoulder Instability Index (WOSI; Shoulder)

The Western Ontario Shoulder Instability Index (WOSI) was developed in 1998 for use in patients with complaints related to shoulder instability. It consists of 21 questions that evaluate physical symptoms, sports/recreation/work, lifestyle, and emotions. Each question is scored on a 100-mm visual analog scale (VAS), making the total score 2100. Lower scores indicate less impairment and symptoms [40, 66]. Because this instrument was designed with a rather narrow focus (for use in shoulder instability only), it is reasonable to report the minimal clinically important difference (a concept that will be discussed shortly), which has been reported as 220 (10.4%) in the literature [39].

12.5.6 Oxford Shoulder Score (OSS; Shoulder)

The Oxford Shoulder Score (OSS) is a 12-item PROM developed in 1996 by a group of researchers at Oxford University with the intention of measuring the effect of surgical intervention for a variety of shoulder diagnoses, except instability [17]. Each item is scored on a five-point scale. The sum of the 12 questions is then converted such that higher scores indicate improved outcomes [18]. The OSS has been shown to be valid, reliable, and sensitive to changes over time after shoulder surgery [79].

12.5.7 Hip Disability and Osteoarthritis Outcome Score (HOOS) and HOOS Joint Replacement (HOOS JR; Hip)

The Hip Disability and Osteoarthritis Outcome Survey (HOOS) is a 40-item outcome measure developed by Roos and colleagues in the early 2000s. It includes the Western Ontario McMaster osteoarthritis score (WOMAC) in its entirety and adds an additional two dimensions that assess sports/recreation and hip-related quality of life, thus allowing the HOOS to capture five total domains (including pain, other symptoms, and activities of daily living, which are captured by the WOMAC) [52].

One aspect that makes the HOOS (and the KOOS) unique is that a total score is not calculated. Rather than combined scores, each domain is scored separately, producing five subscale scores that range from 0 to 100, with higher scores indicating better outcomes. This is both a strength and limitation to the HOOS (and KOOS), as the subscales allow for more granular assessment of changes to particular domains, while the inability to calculate a total score leads to more significant floor/ceiling effects and limits one's ability to compare HOOS scores to other PROMs that generate a single score. The HOOS has been shown to have excellent psychometric properties [41]. Another advantage of the HOOS is that it is responsive enough for use in both short- and long-term studies.

The HOOS JR is a six-question short-form version of the HOOS that was developed by Lyman and colleagues at the Hospital for Special Surgery. It utilizes questions from the "pain" and "activities of daily living" domains and has been validated for use in patients undergoing total hip arthroplasty (THA) [46].

12.5.8 Knee Injury and Osteoarthritis Outcome Score (KOOS) and KOOS Joint Replacement (KOOS JR; Knee)

Like the HOOS, the Knee injury and Osteoarthritis Outcome Score (KOOS) was developed by Roos et al. in the late 1990s. It too includes the

WOMAC in its entirety and thereafter adds the same two domains that were developed to create the HOOS (sports/recreation and knee-related quality of life) [61]. In some ways, the inclusion of the WOMAC represents a limitation of the KOOS (and the HOOS), as two of the most commonly performed knee procedures—anterior cruciate ligament reconstruction and arthroscopic partial meniscectomy—are typically done on younger populations that do not have osteoarthritis. However, the benefit of having the WOMAC built into the questionnaire is that the KOOS (or the HOOS) can be used in prospective cohort studies or longitudinal databases involving populations in which osteoarthritis (whether primary or posttraumatic) is expected to occur.

The KOOS has been validated for use in a variety of knee diagnosis ranging from osteoarthritis to ACL rupture [5, 22, 36]. The KOOS subscales are also kept separate, although a "global" score, dubbed the KOOS4, has been utilized in large randomized trials (despite a lack of evidence supporting the use of the KOOS subscales in this fashion) [24]. The KOOS4 represents an average of four KOOS subscales (likewise, the KOOS5 represents an average of all five KOOS subscales) and is useful for assessing outcomes for interventions that are expected to affect multiple KOOS subscales simultaneously. The KOOS has been demonstrated to have excellent psychometric properties as well [15]. Like the HOOS, the KOOS is suitable for use in both short- and long-term studies.

The KOOS JR is a seven-question short-form version of the KOOS that was developed by the same HSS group that pioneered the HOOS JR and has been validated for use in total knee arthroplasty (TKA) [45].

12.5.9 International Knee Documentation Committee Subjective Knee Form (IKDC-SKF; Knee)

The International Knee Documentation Committee Subjective Knee Form (IKDC-SKF) debuted in the early 2000s, after years of devel-

opment by Irrgang et al., and features ten questions that address symptoms, sports activities, and function [34]. It was intended to be a PROM that could be used for any condition involving the knee and has been shown to be valid, reliable, and responsive across multiple diagnoses [26, 30, 35]. The raw score is transformed to a 100-point scale, with higher scores indicating better outcomes.

12.5.10 Marx Activity Rating Scale (Knee)

The Marx Activity Rating Scale was developed by Marx and colleagues and debuted in 2001. It is a four-item questionnaire that assesses the peak frequency of patients' participation in running, cutting, deceleration, and pivoting activities over the last 12 months [49]. Answers are on a 0–4 scale, with a maximum score of 16. Higher scores indicate more frequent participation in the aforementioned activities.

12.5.11 Tegner Activity Level (Knee)

The Tegner Activity Scale is an alternative to the Marx. It asks patients to indicate the highest level of activity that they participated in prior to their knee injury and the highest level that they are

Clinical Vignette 3

One of the largest and most well-known studies in sports medicine is the Multicenter Orthopaedics Outcome Network (MOON) ACL reconstruction prospective cohort study. The MOON cohort now includes nearly 3500 ACL reconstructions and has achieved >80% follow-up at 2, 6, and 10 years post-op. The MOON group was one of the first to use PROMs as a primary outcome in an ACL reconstruction cohort. The co-investigators of MOON chose to utilize three PROMs—KOOS, IKDC, and Marx [67].

currently able to participate. Scores range from 0 (unable to work because of knee problems) to 10 (competitive sports such as soccer, football, rugby on a national/elite level) [10, 72].

12.5.12 Foot and Ankle Ability Measure (FAAM; Foot and Ankle)

The Foot and Ankle Ability Measure (FAAM) is a region-specific PROM designed by Martin et al. in 2005 for use in patients with leg, ankle, and foot disorders. It contains two subscales—“sports” and “activities of daily living”—which, combined, have 29 questions that are scored on a five-point Likert scale. The raw scores of each subscale are kept separate and converted to a percentage, with higher scores indicating better outcomes. The FAAM has been shown to be valid, reliable, and responsive across a spectrum of diagnoses involving the leg, foot, and ankle [12, 20, 47, 48].

12.5.13 Foot and Ankle Disability Index (FADI; Foot and Ankle)

The Foot and Ankle Disability Index (FADI) was also designed by Martin et al. and includes the FAAM in its entirety along with an additional five questions (four regarding pain and one regarding sleep). It is scored in the same fashion as the FAAM and interpreted in the same way as well. Like the FAAM, the FADI has been shown to have acceptable psychometric properties [20, 28].

12.5.14 Foot and Ankle Outcome Score (FAOS; Foot and Ankle)

The Foot and Ankle Outcome Score (FAOS) is a lower leg adaptation of the KOOS and features the same number of questions in the same subscale format [60]. Like the KOOS, the five FAOS subscales are scored separately and range from 0 to 100, with higher scores indicating superior outcomes.

12.5.15 Achilles Tendon Total Rupture Score (ATRS; Foot and Ankle)

The Achilles Tendon Total Rupture Score (ATRS) was developed to assess outcomes of patients having suffered total Achilles tendon ruptures. It features ten questions and has shown satisfactory psychometric properties [53].

12.6 Single-Item Measures of Outcome

While PROMs such as those mentioned above provide significant detail in terms of symptoms and function from the patient's perspective, they are often difficult to interpret clinically, particularly at baseline. Single-item outcome measures such as the Single Assessment Numeric Evaluation (SANE) and the Patient Acceptable Symptom State (PASS) are more readily interpreted in the clinic, are very brief to administer, and can provide additional information when used in conjunction with lengthier PROMs.

12.6.1 Single Assessment Numeric Evaluation (SANE)

The Single Assessment Numeric Evaluation (SANE) is a single question that asks, "How would you rate your [joint] today as a percentage of normal (0 to 100% scale, with 100% being normal)?" Despite its simplicity, the SANE correlates well with longer PROMs such as the IKDC (for use in knee-related diagnoses), the ASES (for use in shoulder-related diagnoses), and other PROMs [56, 57, 65, 74, 75, 77]. The advantages of the SANE include its brevity (one question) and interpretability (from the patients' perspectives).

However, the major shortcoming of the SANE (and other single-item outcome measures) is its multidimensional nature, which inhibits interpretability when patients report low SANE scores, as the clinician or researcher is unable to

identify the driver of the poor outcome (e.g., the patient could have pain, poor function, additional symptoms, or a combination of the three).

The SANE is typically administered as an adjunct to other PROMs such as the IKDC or the ASES rather than as a standalone PROM.

12.6.2 Patient Acceptable Symptom State (PASS)

The Patient Acceptable Symptom State (PASS) is another single-item outcome measure that has grown in popularity over the past decade. The PASS is believed to be a threshold beyond which the patient "feels well" [43]. Although there are multiple versions of the PASS question, one of the most common versions is "taking into account all the activity you have during your daily life, your level of pain, and also your activity limitations and participation restrictions, do you consider the current state of your [joint] satisfactory?" Much of the current research involving the PASS aims to identify thresholds of other commonly used PROMs (such as the KOOS or the IKDC) that correlate with PASS [21, 33, 51, 76]. It remains unclear whether achievement of PASS is driven by a degree of change in symptoms or by the realization of a symptom threshold. This is an area currently under investigation.

Like the SANE, the PASS is typically administered alongside other PROMs rather than as a standalone measure.

12.7 Observer-Reported Outcome Measures (OROMs)

Patient-reported outcome measures provide a wealth of information regarding the impact of disease on patient symptoms, quality of life, and function from the patient's perspective. However, they fail to capture certain variables that are more relevant to physicians than to patients (e.g., laxity measurement following an ACL reconstruction). Thus, observer-reported

outcome measures (OROMs) are also important tools for use in clinical research. These include physical exam measurements, functional tests, and imaging classification schemes. The list of available OROMs is even longer than the list of commonly used PROMs, and, thus, this portion of the chapter will review important characteristics of OROMs to consider when choosing one rather than attempting to review a handful of OROMs.

12.7.1 Physical Examination Measurements

Physical exam measurements can be separated into two major categories—traditional “hands-on” techniques and instrumented techniques.

12.7.2 Traditional “Hands-On” Techniques

Traditional “hands-on” techniques include basic physical exam variables such as range of motion (ROM) and strength, as well as more nuanced examination techniques or special tests (e.g., Lachman exam, meniscal testing, special examination maneuvers of the shoulder, etc.). When choosing which, if any, of these OROMs to include when designing a study, it is important to consider characteristics such as reliability (both inter and intra-rater), repeatability, and reproducibility [3].

12.7.3 Instrumented Techniques

Instrumented techniques, such as the use of the KT-1000 arthrometer or the Telos SD 900 for the measurement of knee laxity, are used to minimize measurement error (thus increasing the reliability of a measurement) [8]. While the increase in reliability is beneficial, instrumented techniques typically increase the cost of conducting a study considerably when compared to using “hands-on” techniques.

12.7.4 Functional Tests

Another category of OROMs commonly used in orthopedic research is functional testing. One example of a commonly used functional test is the hop test following ACL reconstruction [1]. In an ideal world, a functional test bridges the gap between a PROM and a pure OROM; however, functional tests tend to be limited by significant issues in terms of reliability and reproducibility.

12.7.5 Imaging Measurement Techniques

Imaging measurement techniques are other commonly used OROMs. Examples of such measures include the Kellgren-Lawrence grading scheme and the Osteoarthritis Research Society International (OARSI) classification score, both of which are used to assess radiographic changes associated with knee osteoarthritis [42]. Additional examples include the modified Outerbridge scale, the Whole-Organ Magnetic Resonance Imaging Score (WORMS), and the Knee Osteoarthritis Scoring System (KOSS), all of which can be used to evaluate magnetic resonance images in a way that is standardized and reliable [55, 78].

Take-Home Message

- Numerous scoring instruments are available for use in clinical research.
- It is important to recognize the strengths and limitations of the instrument(s) that will be used when designing a clinical research study.
- No perfect scoring instrument exists, and the soundest clinical research typically utilizes a combination of multiple instruments including both PROMs (general and joint-specific) and OROMs.

References

1. Abrams GD, Harris JD, Gupta AK, et al. Functional performance testing after anterior cruciate ligament reconstruction. *Orthop J Sports Med.* 2014;2(1):2325967113518305. <https://doi.org/10.1177/2325967113518305>.

2. Angst F, Schwyzer H-K, Aeschlimann A, Simmen BR, Goldhahn J. Measures of adult shoulder function: Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH) and its short version (QuickDASH), Shoulder Pain and Disability Index (SPADI), American Shoulder and Elbow Surgeons (ASES) Society Standardized Shoulder Assessment Form, Constant (Murley) Score (CS), Simple Shoulder Test (SST), Oxford Shoulder Score (OSS), Shoulder Disability Questionnaire (SDQ), and Western Ontario Shoulder Instability Index (WOSI). *Arthritis Care Res.* 2011;63(S11):S174–88. <https://doi.org/10.1002/acr.20630>.
3. Bartlett JW, Frost C. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet Gynecol.* 2008;31(4):466–75. <https://doi.org/10.1002/uog.5256>.
4. Beaton DE, Wright JG, Katz JN, Upper Extremity Collaborative Group. Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am.* 2005;87(5):1038–46. <https://doi.org/10.2106/JBJS.D.02060>.
5. Bekkers JEJ, de Windt TS, Raijmakers NJH, Dhert WJA, Saris DBF. Validation of the Knee Injury and Osteoarthritis Outcome Score (KOOS) for the treatment of focal cartilage lesions. *Osteoarthr Cartil.* 2009;17(11):1434–9. <https://doi.org/10.1016/j.joca.2009.04.019>.
6. Berlowitz DR, Foy CG, Kazis LE, et al. Effect of intensive blood-pressure treatment on patient-reported outcomes. *N Engl J Med.* 2017;377(8):733–44. <https://doi.org/10.1056/NEJMoa1611179>.
7. Bhatt A. Evolution of clinical research: a history before and beyond James Lind. *Perspect Clin Res.* 2010;1(1):6–10.
8. Boyer P, Djian P, Christel P, Paoletti X, Degeorges R. [Reliability of the KT-1000 arthrometer (Medmetric) for measuring anterior knee laxity: comparison with Telos in 147 knees]. *Rev Clin Orthop Reparatrice Appar Mot.* 2004;90(8):757–64.
9. Brand RA. Ernest Amory Codman, MD, 1869–1940. *Clin Orthop.* 2009;467(11):2763–5. <https://doi.org/10.1007/s11999-009-1047-8>.
10. Briggs KK, Steadman JR, Hay CJ, Hines SL. Lysholm score and Tegner activity level in individuals with normal knees. *Am J Sports Med.* 2009;37(5):898–901. <https://doi.org/10.1177/0363546508330149>.
11. Browne JP, Cano SJ, Smith S. Using patient-reported outcome measures to improve health care: time for a new approach. *Med Care.* 2017;55(10):901. <https://doi.org/10.1097/MLR.0000000000000792>.
12. Garcia CR, Martin RL, Drouin JM. Validity of the foot and ankle ability measure in athletes with chronic ankle instability. *J Athl Train.* 2008;43(2):179–83.
13. Codman EA. The classic: the registry of bone sarcomas as an example of the end-result idea in hospital organization. *Clin Orthop.* 2009;467(11):2766–70. <https://doi.org/10.1007/s11999-009-1048-7>.
14. Collier R. Legumes, lemons and streptomycin: a short history of the clinical trial. *Can Med Assoc J.* 2009;180(1):23–4. <https://doi.org/10.1503/cmaj.081879>.
15. Collins NJ, Prinsen CA, Christensen R, Bartels EM, Terwee CB, Roos EM. Knee Injury and Osteoarthritis Outcome Score (KOOS): systematic review and meta-analysis of measurement properties. *Osteoarthr Cartil.* 2016;24(8):1317–29. <https://doi.org/10.1016/j.joca.2016.03.010>.
16. Davis JC, Bryan S. Patient Reported Outcome Measures (PROMs) have arrived in sports and exercise medicine: why do they matter? *Br J Sports Med.* 2015;49(24):1545–6. <https://doi.org/10.1136/bjsports-2014-093707>.
17. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg Br.* 1996;78(4):593–600.
18. Dawson J, Rogers K, Fitzpatrick R, Carr A. The Oxford shoulder score revisited. *Arch Orthop Trauma Surg.* 2009;129(1):119–23. <https://doi.org/10.1007/s00402-007-0549-7>.
19. Devlin NJ, Brooks R. EQ-5D and the EuroQol group: past, present and future. *Appl Health Econ Health Policy.* 2017;15(2):127–37. <https://doi.org/10.1007/s40258-017-0310-5>.
20. Eecheute C, Vaes P, Van Aerschot L, Asman S, Duquet W. The clinimetric qualities of patient-assessed instruments for measuring chronic ankle instability: a systematic review. *BMC Musculoskelet Disord.* 2007;8:6. <https://doi.org/10.1186/1471-2474-8-6>.
21. Emerson Kavchak AJ, Cook C, Hegedus EJ, Wright AA. Identification of cut-points in commonly used hip osteoarthritis-related outcome measures that define the patient acceptable symptom state (PASS). *Rheumatol Int.* 2013;33(11):2773–82. <https://doi.org/10.1007/s00296-013-2813-1>.
22. Engelhart L, Nelson L, Lewis S, et al. Validation of the knee injury and osteoarthritis outcome score subscales for patients with articular cartilage lesions of the knee. *Am J Sports Med.* 2012;40(10):2264–72. <https://doi.org/10.1177/0363546512457646>.
23. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy.* 1980;66(8):271–3.
24. Frobell RB, Roos EM, Roos HP, Ranstam J, Lohmander LS. A randomized trial of treatment for acute anterior cruciate ligament tears. *N Engl J Med.* 2010;363(4):331–42. <https://doi.org/10.1056/NEJMoa0907797>.
25. Garratt A, Schmidt L, Mackintosh A, Fitzpatrick R. Quality of life measurement: bibliographic study of patient assessed health outcome measures. *BMJ.* 2002;324(7351):1417.
26. Greco NJ, Anderson AF, Mann BJ, et al. Responsiveness of the International Knee Documentation Committee Subjective Knee Form in comparison to the Western Ontario and McMaster Universities Osteoarthritis Index, modified Cincinnati Knee Rating System, and Short Form

- 36 in patients with focal articular cartilage defects. *Am J Sports Med.* 2010;38(5):891–902. <https://doi.org/10.1177/0363546509354163>.
27. Hains F, Waalen J, Mior S. Psychometric properties of the neck disability index. *J Manip Physiol Ther.* 1998;21(2):75–80.
 28. Hale SA, Hertel J. Reliability and sensitivity of the foot and ankle disability index in subjects with chronic ankle instability. *J Athl Train.* 2005;40(1):35–40.
 29. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Qual Life Res.* 2009;18(7):873–80. <https://doi.org/10.1007/s1136-009-9496-9>.
 30. Higgins LD, Taylor MK, Park D, et al. Reliability and validity of the International Knee Documentation Committee (IKDC) Subjective Knee Form. *Joint Bone Spine.* 2007;74(6):594–9. <https://doi.org/10.1016/j.jbspin.2007.01.036>.
 31. Hu Y, Lv G, Ren S, Johansen D. Mid- to long-term outcomes of cervical disc arthroplasty versus anterior cervical discectomy and fusion for treatment of symptomatic cervical disc disease: a systematic review and meta-analysis of eight prospective randomized controlled trials. *PLoS One.* 2016;11(2):e0149312. <https://doi.org/10.1371/journal.pone.0149312>.
 32. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med.* 1996;29(6):602–8. [https://doi.org/10.1002/\(SICI\)1097-0274\(199606\)29:6<602::AID-AJIM4>3.0.CO;2-L](https://doi.org/10.1002/(SICI)1097-0274(199606)29:6<602::AID-AJIM4>3.0.CO;2-L).
 33. Ingelsrud LH, Granan L-P, Terwee CB, Engebretsen L, Roos EM. Proportion of patients reporting acceptable symptoms or treatment failure and their associated KOOS values at 6 to 24 months after anterior cruciate ligament reconstruction: a study from the Norwegian Knee Ligament Registry. *Am J Sports Med.* 2015;43(8):1902–7. <https://doi.org/10.1177/0363546515584041>.
 34. Irrgang JJ, Anderson AF, Boland AL, et al. Development and validation of the international knee documentation committee subjective knee form. *Am J Sports Med.* 2001;29(5):600–13. <https://doi.org/10.1177/03635465010290051301>.
 35. Irrgang JJ, Anderson AF, Boland AL, et al. Responsiveness of the International Knee Documentation Committee Subjective Knee Form. *Am J Sports Med.* 2006;34(10):1567–73. <https://doi.org/10.1177/0363546506288855>.
 36. Johnson DS, Smith RB. Outcome measurement in the ACL deficient knee—what’s the score? *Knee.* 2001;8(1):51–7.
 37. Jones D, Kazis L, Lee A, et al. Health status assessments using the veterans SF-12 and SF-36: methods for evaluating outcomes in the Veterans Health Administration. *J Ambul Care Manage.* 2001;24(3):68–86.
 38. Kazis LE, Ren XS, Lee A, et al. Health status in VA patients: results from the Veterans Health Study. *Am J Med Qual.* 1999;14(1):28–38. <https://doi.org/10.1177/106286069901400105>.
 39. Kirkley A, Griffin S, Dainty K. Scoring systems for the functional assessment of the shoulder. *Arthroscopy.* 2003;19(10):1109–20. <https://doi.org/10.1016/j.arthro.2003.10.030>.
 40. Kirkley A, Griffin S, McLintock H, Ng L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability. The Western Ontario Shoulder Instability Index (WOSI). *Am J Sports Med.* 1998;26(6):764–72. <https://doi.org/10.1177/03635465980260060501>.
 41. Klässbo M, Larsson E, Mannevik E. Hip disability and osteoarthritis outcome score. An extension of the Western Ontario and McMaster Universities Osteoarthritis Index. *Scand J Rheumatol.* 2003;32(1):46–51.
 42. Kohn MD, Sassoon AA, Fernando ND. Classifications in brief: Kellgren-Lawrence classification of osteoarthritis. *Clin Orthop.* 2016;474(8):1886–93. <https://doi.org/10.1007/s11999-016-4732-4>.
 43. Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? *Ann Rheum Dis.* 2007;66(Suppl 3):iii40–1. <https://doi.org/10.1136/ard.2007.079798>.
 44. Laucis NC, Hays RD, Bhattacharyya T. Scoring the SF-36 in orthopaedics: a brief guide. *J Bone Joint Surg Am.* 2015;97(19):1628–34. <https://doi.org/10.2106/JBJS.O.00030>.
 45. Lyman S, Lee Y-Y, Franklin PD, Li W, Cross MB, Padgett DE. Validation of the KOOS, JR: a short-form knee arthroplasty outcomes survey. *Clin Orthop.* 2016;474(6):1461–71. <https://doi.org/10.1007/s11999-016-4719-1>.
 46. Lyman S, Lee Y-Y, Franklin PD, Li W, Mayman DJ, Padgett DE. Validation of the HOOS, JR: a short-form hip replacement survey. *Clin Orthop.* 2016;474(6):1472–82. <https://doi.org/10.1007/s11999-016-4718-2>.
 47. Martin RL, Irrgang JJ. A survey of self-reported outcome instruments for the foot and ankle. *J Orthop Sports Phys Ther.* 2007;37(2):72–84. <https://doi.org/10.2519/jospt.2007.2403>.
 48. Martin RL, Irrgang JJ, Burdett RG, Conti SF, Van Swearingen JM. Evidence of validity for the foot and ankle ability measure (FAAM). *Foot Ankle Int.* 2005;26(11):968–83. <https://doi.org/10.1177/107110070502601113>.
 49. Marx RG, Stump TJ, Jones EC, Wickiewicz TL, Warren RF. Development and evaluation of an activity rating scale for disorders of the knee. *Am J Sports Med.* 2001;29(2):213–8.
 50. Michener LA, McClure PW, Sennett BJ. American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report section: reliability, validity, and responsiveness. *J Shoulder Elb*

- Surg. 2002;11(6):587–94. <https://doi.org/10.1067/mse.2002.127096>.
51. Muller B, Yabroudi MA, Lynch A, et al. Defining thresholds for the patient acceptable symptom state for the IKDC Subjective Knee Form and KOOS for patients who underwent ACL reconstruction. *Am J Sports Med.* 2016;44(11):2820–6. <https://doi.org/10.1177/0363546516652888>.
 52. Nilsdotter AK, Lohmander LS, Klässbo M, Roos EM. Hip disability and osteoarthritis outcome score (HOOS)—validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord.* 2003;4:10. <https://doi.org/10.1186/1471-2474-4-10>.
 53. Nilsson-Helander K, Thomeé R, Silbernagel KG, et al. The Achilles tendon Total Rupture Score (ATRS): development and validation. *Am J Sports Med.* 2007;35(3):421–6. <https://doi.org/10.1177/0363546506294856>.
 54. Oak SR, Strnad GJ, Bena J, et al. Responsiveness comparison of the EQ-5D, PROMIS Global Health, and VR-12 questionnaires in knee arthroscopy. *Orthop J Sports Med.* 2016;4(12):2325967116674714. <https://doi.org/10.1177/2325967116674714>.
 55. Peterfy CG, Guermazi A, Zaim S, et al. Whole-organ magnetic resonance imaging score (WORMS) of the knee in osteoarthritis. *Osteoarthr Cartil.* 2004;12(3):177–90. <https://doi.org/10.1016/j.joca.2003.11.003>.
 56. Pietrosimone B, Luc BA, Duncan A, Saliba SA, Hart JM, Ingersoll CD. Association between the single assessment numeric evaluation and the Western Ontario and McMaster Universities Osteoarthritis Index. *J Athl Train.* 2017;52(6):526–33. <https://doi.org/10.4085/1062-6050-52.5.07>.
 57. Provencher MT, Frank RM, Macian D, et al. An analysis of shoulder outcomes scores in 275 consecutive patients: disease-specific correlation across multiple shoulder conditions. *Mil Med.* 2012;177(8):975–82. <https://doi.org/10.7205/MILMED-D-11-00234>.
 58. Radcliff K, Davis RJ, Hisey MS, et al. Long-term evaluation of cervical disc arthroplasty with the Mobi-C® cervical disc: a randomized, prospective, multicenter clinical trial with seven-year follow-up. *Int J Spine Surg.* 2017;11:31. <https://doi.org/10.14444/4031>.
 59. Richards RR, An KN, Bigliani LU, et al. A standardized method for the assessment of shoulder function. *J Shoulder Elb Surg.* 1994;3(6):347–52. [https://doi.org/10.1016/S1058-2746\(09\)80019-0](https://doi.org/10.1016/S1058-2746(09)80019-0).
 60. Roos EM, Brandsson S, Karlsson J. Validation of the foot and ankle outcome score for ankle ligament reconstruction. *Foot Ankle Int.* 2001;22(10):788–94. <https://doi.org/10.1177/107110070102201004>.
 61. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. *J Orthop Sports Phys Ther.* 1998;28(2):88–96. <https://doi.org/10.2519/jospt.1998.28.2.88>.
 62. Samuelsen BT, Webster KE, Johnson NR, Hewett TE, Krych AJ. Hamstring autograft versus patellar tendon autograft for ACL reconstruction: is there a difference in graft failure rate? A meta-analysis of 47,613 patients. *Clin Orthop.* 2017;475(10):2459–68. <https://doi.org/10.1007/s11999-017-5278-9>.
 63. Schalet BD, Rothrock NE, Hays RD, et al. Linking physical and mental health summary scores from the veterans RAND 12-Item Health Survey (VR-12) to the PROMIS® Global Health Scale. *J Gen Intern Med.* 2015;30(10):1524–30. <https://doi.org/10.1007/s11606-015-3453-9>.
 64. Schmidt S, Franke J, Rauschmann M, Adelt D, Bonsanto MM, Sola S. Prospective, randomized, multicenter study with 2-year follow-up to compare the performance of decompression with and without interlaminar stabilization. *J Neurosurg Spine.* 2018;28:1–10. <https://doi.org/10.3171/2017.1.SPINE17643>.
 65. Shelbourne KD, Barnes AF, Gray T. Correlation of a single assessment numeric evaluation (SANE) rating with modified Cincinnati knee rating system and IKDC subjective total scores for patients after ACL reconstruction or knee arthroscopy. *Am J Sports Med.* 2012;40(11):2487–91. <https://doi.org/10.1177/0363546512458576>.
 66. Smith MV, Calfee RP, Baumgarten KM, Brophy RH, Wright RW. Upper extremity-specific measures of disability and outcomes in orthopaedic surgery. *J Bone Joint Surg Am.* 2012;94(3):277–85. <https://doi.org/10.2106/JBJS.J.01744>.
 67. Spindler KP, Huston LJ. *O'Donoghue Sports Injury Award* 10 year outcomes and risk factors after ACL reconstruction: a multicenter cohort study. *Orthop J Sports Med.* 2017;5(7 Suppl 6):2325967117S00247. <https://doi.org/10.1177/2325967117S00247>.
 68. Spindler KP, Kuhn JE, Freedman KB, Matthews CE, Dittus RS, Harrell FE. Anterior cruciate ligament reconstruction autograft choice: bone-tendon-bone versus hamstring: does it really matter? A systematic review. *Am J Sports Med.* 2004;32(8):1986–95.
 69. Sur RL, Dahm P. History of evidence-based medicine. *Indian J Urol.* 2011;27(4):487–9. <https://doi.org/10.4103/0970-1591.91438>.
 70. Svensson M, Sernert N, Ejerhed L, Karlsson J, Kartus JT. A prospective comparison of bone-patellar tendon-bone and hamstring grafts for anterior cruciate ligament reconstruction in female patients. *Knee Surg Sports Traumatol Arthrosc.* 2006;14(3):278–86. <https://doi.org/10.1007/s00167-005-0708-8>.
 71. Tarlov AR, Ware JE, Greenfield S, Nelson EC, Perrin E, Zubkoff M. The Medical Outcomes Study. An application of methods for monitoring the results of medical care. *JAMA.* 1989;262(7):925–30.
 72. Tegner Y, Lysholm J. Rating systems in the evaluation of knee ligament injuries. *Clin Orthop.* 1985;198:43–9.
 73. Vernon H, Mior S. The Neck Disability Index: a study of reliability and validity. *J Manip Physiol Ther.* 1991;14(7):409–15.
 74. Williams GN, Gangel TJ, Arciero RA, Uhorchak JM, Taylor DC. Comparison of the single assessment

- numeric evaluation method and two shoulder rating scales. Outcomes measures after shoulder surgery. *Am J Sports Med.* 1999;27(2):214–21. <https://doi.org/10.1177/03635465990270021701>.
75. Williams GN, Taylor DC, Gangel TJ, Uhorchak JM, Arciero RA. Comparison of the single assessment numeric evaluation method and the Lysholm score. *Clin Orthop.* 2000;373:184–92.
76. Wright AA, Hensley CP, Gilbertson J, Leland JM III, Jackson S. Defining patient acceptable symptom state thresholds for commonly used patient reported outcomes measures in general orthopedic practice. *Man Ther.* 2015;20(6):814–9. <https://doi.org/10.1016/j.math.2015.03.011>.
77. Wright RW, Baumgarten KM. Shoulder outcomes measures. *J Am Acad Orthop Surg.* 2010;18(7):436–44.
78. Wright RW, Ross JR, Haas AK, et al. Osteoarthritis classification scales: interobserver reliability and arthroscopic correlation. *J Bone Joint Surg Am.* 2014;96(14):1145–51. <https://doi.org/10.2106/JBJS.M.00929>.
79. Younis F, Sultan J, Dix S, Hughes PJ. The range of the Oxford Shoulder Score in the asymptomatic population: a marker for post-operative improvement. *Ann R Coll Surg Engl.* 2011;93(8):629–33. <https://doi.org/10.1308/003588411X13165261994193>.
80. Zaina F, Tomkins-Lane C, Carragee E, Negrini S. Surgical versus non-surgical treatment for lumbar spinal stenosis. *Cochrane Database Syst Rev.* 2016;(1):CD010264. <https://doi.org/10.1002/14651858.CD010264.pub2>.
81. Zimmerman AL. Evidence-based medicine: a short history of a modern medical movement. *Virtual Mentor.* 2013;15(1):71. <https://doi.org/10.1001/virtualmentor.2013.15.1.mhst1-1301>.



Health Measurement Development and Interpretation

13

Andrew Firth, Dianne Bryant, Jacques Menetrey,
and Alan Getgood

13.1 What Is an Outcome Measure, and Why Measure Health Outcomes?

In 1948, the World Health Organization defined health as ‘a state of complete physical, mental, and social well-being [28]’. This idealistic concept of health as a comprehensive, theoretical framework is problematic for those attempting to quantify health within an individual to inform decision-making or within or between groups to inform broader clinical recommendations. Despite the complexity of health measurement, outcome measures are designed with one of three purposes in mind: *discrimination*, *evaluation*, or *prediction*. In addition, outcome measures may be *surrogates* for important outcomes or themselves measure outcomes that are of *direct importance to patients*.

A. Firth · A. Getgood (✉)
Fowler Kennedy Sport Medicine Clinic, 3M Centre,
University of Western Ontario, London, ON, Canada
e-mail: afirth5@uwo.ca; alan.getgood@uwo.ca

D. Bryant
Faculty of Health Sciences, Elborn College,
University of Western Ontario, London, ON, Canada
e-mail: dianne.bryant@uwo.ca

J. Menetrey
Centre de médecine du sport et de l'exercice,
Hirslanden Clinique La Colline, University Hospital
of Geneva, Geneva, Switzerland
e-mail: Jacques.menetrey@lacolline.ch

13.1.1 Discriminative, Predictive, and Evaluative Outcomes

To select an appropriate outcome measure for research, the objective of the measurement must be clear and aligned with the objectives of the study, the population being investigated, and the study methodology [7]. For example, an instrument used to *discriminate* between individuals might be used to determine a patient's eligibility for study participation or may be a diagnostic tool designed to classify individuals as negative or positive for disease. For example, the Kellgren and Lawrence system is used to classify the severity of radiographic knee osteoarthritis (OA) [10] and is used in conjunction with clinical symptoms to discriminate between individuals with advanced OA and those with mild OA.

A *predictive* instrument, on the other hand, is used to predict future events. For example, there is some evidence that a test of a patient's landing biomechanics post-anterior cruciate ligament (ACL) reconstruction can predict reinjury [6, 18] and is therefore used by some clinicians as a guide for making recommendations regarding return to sport.

Finally, an *evaluative* measurement tool is used to assess change and is most commonly used to measure the effect of a therapy. For example, if an investigator wishes to determine whether a regimen of aerobic exercise will improve the quality of life, pain, and mobility in

patients with osteoarthritis of the knee, she/he will administer each of the evaluative measures before the intervention and then again after the intervention and assess whether scores have improved. Similarly, if a researcher's objective is to compare the effectiveness of two or more therapies, each group of patients will complete the evaluative measure at the endpoint of interest so that between-group comparisons can be made.

13.1.2 Surrogate- and Patient-Important Outcomes

Surrogate outcomes are not necessarily meaningful to patients but are believed to proxy outcomes that are directly important to patients [2]. Early orthopaedic research often relied solely on surrogate outcomes like imaging, performance-based tests, or physiological tests to provide evidence of a treatment's success. Today, surrogate outcomes are most appropriate for smaller, explanatory, or proof-of-concept studies prior to investing in larger, more pragmatic studies that will apply to broader populations. For example, shoulder range of motion is a proxy for shoulder function but is not a direct measure of function since an individual may not require full motion to complete desired tasks or can find a way to compensate for a loss of motion and still accomplish desired tasks in a fulfilling way. Other examples of common surrogate outcomes are radiographic imaging of joint space narrowing, a surrogate for pain and impaired function; bone mineral density, a surrogate for risk of fragility fracture; strength, a surrogate for functional ability; etc.

Although surrogate measures are important determinants of health, and help to provide explanations for impairments or predict future patient-important health issues, pragmatic studies that aim to make recommendations for practice change should be measuring effectiveness using a measure of direct importance to patients, like

the incidence of adverse outcomes (e.g. death, MI, stroke, revision surgery, etc.) and patient-reported outcome measures (PROMs).

A PROM is a subjective assessment completed by the patient where they are asked to score aspects of their own perception of their health [20]. Some PROMs include only questions related to functional ability, called patient-reported functional ability questionnaires (e.g. Lower Extremity Functional Score (LEFS), American Shoulder and Elbow Surgeons (ASES) Score), while others attempt to measure health from a more comprehensive perspective and include items querying physical, mental, and social well-being (e.g. 12-Item Short-Form Survey (SF-12), Western Ontario Rotator Cuff (WORC) Index). The latter are referred to as health-related quality-of-life (HRQOL) instruments.

HRQOL measures often fall into one of two categories: *disease-specific* or *generic*. Disease-specific measures are designed to ask patients about constructs of health directly affected by the disease in question (e.g. the Western Ontario Rotator Cuff (WORC) Index is specific to patients with rotator cuff pathology). Generic measures, on the other hand, ask questions about health from a non-specific, highly applicable perspective. Thus, while disease-specific measures are more sensitive to changes in health [27], generic measures have a larger scope and can therefore be interpreted across different health states, including those who are healthy [7].

Often, the measurement properties of the most widely accepted outcome measures are extensively published and provide evidence that the measure is accurate, precise, and able to detect change in the population of interest. If this evidence is not available, those hoping to use the instrument may elect to first assess its measurement properties before implementing its use in clinic or as part of a research study or run the risk of collecting uninterpretable data. In the next section, we describe some of the most common measurement properties.

13.2 Measurement Properties of an Outcome Measure

13.2.1 Reliability

Reliability is about the precision of an outcome measure and evaluates whether it produces consistent results when repeatedly administered in a sample from a population with a stable health status [23]. Every outcome score consists of two components: the patient's true score and random error. An outcome measure is considered more reliable when the random error is small, because a patient's test performance will more closely reflect their true score on the outcome measure with less variability over repeated testing.

We can think of reliability in two different ways: *relative* or the extent to which the instrument can differentiate between individuals and *absolute* or the extent to which repeated administrations of the test produce consistent results. The most useful estimate of an instrument's relative reliability is the *standard error of measurement (SEM)*.

The SEM is presented in units consistent with the original tool and is an estimate of the error associated with an individual's score. If the magnitude of the error is known, it can be used to communicate the accuracy of a single patient's score on an outcome measure and can also be used to create a threshold to determine whether a change in score represents real change over and above measurement error. Specifically, the SEM is the square root of the within-patient variance (calculated from the difference in score within a stable group of individuals over repeated measures) for the sample.

Clinical Vignette

Consider the following: your patient is 1-year post ACL reconstruction. One of the criterion that you use to determine readiness to return to sport is whether the patient has a limb symmetry index (LSI) of at least 90% for each of four different hop tests (forward, triple, and crossover hops for distance and 6-m timed hop). Your patient completes the forward hop for distance for both limbs, and

the LSI is 80%. You want to understand the error associated with this measurement. In a study conducted by Reid et al. [19] in 2007 evaluating the measurement properties of the LSI for all four hop tests, they reported the SEM for the LSI for the single leg hop test to be 4.94%. To place a 95% confidence interval around the patient's LSI of 80%, you would multiply the SEM by 1.96 (the corresponding *z*-value for the desired confidence level) to find the lower and upper limit of the possible range of results. In this case, the individual's LSI result would have a 95%CI of $\pm 6.84\%$. In other words, the individual's LSI may be as low as 73% and as high as 87%. As a clinician making decisions about return to sport, you might evaluate the risk of returning to sport if the LSI is actually 73%.

In terms of absolute reliability, the most common methods are test-retest and inter- and intrarater reliability. *Test-retest reliability* assumes that the health of the patient remains stable but that there may be differences in the results related to errors in how the test is performed. For example, the drop vertical jump test requires the individual to drop straight down from a box and upon landing, jump up as high as they can. As you might imagine, even though the status of the individual's knee stability is unlikely to change over three consecutive tests, the individual's landing biomechanics may be quite different each time. Regular instrument calibration is also part of achieving good test-retest reliability.

Fact Box 13.1: Interpreting ICC Scores

ICC scores range from 0 to 1 with values closer to 1 indicating better reliability. One commonly used interpretation of ICC scores was suggested by Shrout and Fleiss in 1979 [22]:

ICC < 0.4 = poor reliability
 0.4 > ICC < 0.75 = fair to good reliability
 ICC > 0.75 = excellent reliability

Intra-rater reliability is an assessment of the agreement between multiple testing sessions scored by the same rater [14]. Intra-rater reliability is an important measurement property when there is some subjectivity about how the test is performed or interpreted. For example, using a goniometer to measure range of motion will carry some error since location of anatomical landmarks and placement of the instrument may vary between repeated administrations. *Inter-rater reliability* measures the agreement between assessments scored by more than one independent rater [14] and is important to understand if more than one person is responsible for measuring patient outcomes. For example, there may be more than one research assistant (RA) working at your clinic who will contribute measurement data to your surveillance database or multiple RAs working at different sites participating in a multi-centre research study. Both intra- and inter-rater reliability are usually improved with standardized protocols and training opportunities for personnel responsible for outcome measurement.

Although there is no consensus regarding the accepted timeframe between testing sessions for any evaluation of reliability, the length of time should be such that true change is unexpected but not so close together that agreement is overestimated because the individual or rater remembers their previous score or rating [1, 15]. Statistics used to express the level of absolute reliability are the *Kappa* for nominal measures (e.g. lift-off test for the presence or absence of a subscapularis tear), a *weighted Kappa* for ordered outcomes (e.g. Kellgren and Lawrence Grade 0, 1, 2, 3, or 4 osteoarthritis), or an *intraclass correlation coefficient* (ICC), for continuous outcome measures like most PROMs.

Although it is important that the results provided by an outcome measure are reproducible, reproducibility does not guarantee validity. For example, a risk assessment for ACL rupture may be highly reliable when completed by a patient with knee osteoarthritis, yet it provides no useful information regarding disease progression as the questionnaire lacks context and meaning in this population. Reliability is important, but we require validity to provide meaning to outcome measures.

13.2.2 Validity

Evaluating the validity of an outcome measure is all about providing evidence that an instrument is able to measure what it was intended to measure [17]. It's important that assessments of validity take place within the population of interest even if there is sufficient evidence of validity in other populations. There are several methods to evaluate validity. In this chapter, we will describe *face, content, criterion, and construct validity*.

Face validity is a superficial and insufficient means to assess validity. The basic premise is that persons with perhaps insufficient expertise and an unsystematic and incomprehensive approach comment as to whether the instrument appears to measure what is intended [16].

On the other hand, when developing an instrument, content validity is approached by systematically operationalizing the content that comprehensively represents the construct. For example, when an instrument is being developed, content validity is likely to be achieved if items are generated by a large sample of individuals who have experience with the disease including patients, their caregivers and family members, and expert clinicians [17]. Post development, proper evaluation of content validity involves large patient groups and expert clinicians who use set criteria to determine whether the construct of health as it relates to a specific health issue has been captured comprehensively. For example, a new questionnaire meant to measure quality of life in patients suffering from pathology related to the foot and ankle may begin by organizing items into broad categories of health (physical, social, mental) and then into domains that might include such topics as pain, functional ability for ADLs, work expectations, participation in family and community roles, anxiety, depression, etc.

Criterion validity compares the results of the new instrument to the best available method (often coined 'gold standard') to evaluate their association to one another; the larger the association, the more evidence there is for criterion validity [17]. For instance, an open surgical procedure was at one time the only means to evaluate the integrity of the rotator cuff. When

advances in imagining became available, like magnetic resonance imaging and ultrasound, these imaging methods were compared to observations during surgery to determine their accuracy. In this example, the observations made during surgery served as the gold standard. Statistics like sensitivity and specificity are often used to communicate accuracy.

True gold standards are common for measures of structures, physiology, physical function, and performance; however, they rarely exist for constructs like HRQOL. For this more abstract paradigm, we evaluate validity using a theoretical framework of health against which we make assumptions about how the instrument should function if it's accurately capturing the construct. This type of validity is referred to as construct validity [9].

To assess construct validity, we hypothesize the magnitude and direction of the correlation between the new outcome measure and related aspects of health to determine whether the new test behaves as expected. For example, we may hypothesize that as the severity of disease increases, the quality of life score will decrease (directional hypothesis) and that the association between severity and quality of life will be small (expected magnitude). As part of constructing these hypotheses, it's common to compare the performance of the new instrument to outcome measures designed to assess a similar attribute (*convergent validity*) [17, 23]. For example, we may hypothesize that as the score on our new disease-specific quality of life questionnaire increases, so too will the physical component score (PCS) of the generic health-related quality of life instrument, the 12-Item Short-Form Survey (SF12).

Conversely, *discriminant validity* assesses whether the outcome measure performs better than an instrument designed to assess a general or unrelated construct when measuring the outcome of interest [17, 23]. For instance, the investigator could compare a new knee OA-specific questionnaire, to a generic measure of mental health, like the mental component score (MCS) of the SF12, and expect that mental health is only very weakly associated to OA-related quality of life.

Both criterion and construct validity can be evaluated at a single point in time as just described, called cross-sectional validity, or, over time, to evaluate whether changes in the new instrument are associated to changes in existing measures [3, 11, 13, 25]. If we return to our example of a new knee OA-specific questionnaire, we could hypothesize that, over time, the mental component score (MCS) of the SF12 may remain relatively unchanged despite small to moderate changes in the knee OA-specific questionnaire. Selecting outcome measures that have demonstrated longitudinal construct validity within the relevant study population is important when designing clinical trials where instruments are used to assess health status prior to and following treatment.

Although important, validity is insufficient for an investigator wanting to make clinical recommendations based on the results of an evaluative outcome measure. While validity shows that the outcome measure is associated with changes in health, it lacks interpretability. Interpretability of changes in health states begins with our final two measurement properties, sensitivity to change and responsiveness. *Sensitivity to change* refers to the ability of an outcome measure to detect changes in health when change has occurred but does not provide any insight as to the significance of the change to those being treated [13]. Conversely, *responsiveness* describes the outcome measure's ability to detect meaningful change from the patient's perspective, no matter how small the change is [13].

13.2.3 Responsiveness

Assessing responsiveness requires a marker of clinical importance. Traditional methods of determining responsiveness use either an anchor-based (patient defined) or distribution-based (statistical) approach. The goal of either method is to define the minimally clinically important difference (MCID). MCID, originally proposed by Jaeschke et al. [8] in 1989, is the smallest change in the outcome of interest that informed parties (patient/clinician) consider important, whether

favourable or harmful, which would cause those involved to consider changing treatment [21].

Using the anchor-based approach requires the patient to complete the new questionnaire before and after change is expected and to complete a global rating of change (GRC) questionnaire. A GRC requires the patient to indicate whether they have improved, remained unchanged, or worsened; and if they have changed, they next indicate by how much they have changed. An estimate of the MCID is made by calculating the average change score of patients who indicated that they changed by a ‘small but important amount’ on the GRC.

When determining responsiveness from a distribution-based perspective, investigators may construct two normal distributions: the first constructed from the change scores of patients who should have remained stable over the two measurements and the second constructed from the change scores of patients who are expected to have changed over the two measurements. The value of the MCID is said to be greater than the majority of scores in the first distribution and less than the majority of scores in the second distribution.

13.3 Interpreting the Results: Bias and Precision

Once the appropriate outcome measures have been used to collect data, the challenge becomes presenting the data in a meaningful way. Simply presenting the results of statistical tests of the data (i.e. p -values) or the summary measures (e.g. mean change pre- to post- or mean difference between groups) is often meaningless to those without an incredible amount of experience using the same outcome measure, which is extremely rare outside of intensive research programmes.

The problem with only presenting p -values (the probability of an observed result given the null hypothesis is true) [26] is that p -values are highly sample-specific because they are influenced by the sample’s variability and sample size. A p -value does not express the magnitude of effect, the reproducibility of the result, or the importance of

the difference (i.e. Is the difference sufficient to change practice?). Specifically, a study result may achieve statistical significance even though the difference is small and unimportant if the sample is highly homogeneous and/or quite large. On the other hand, an important difference may never reach statistical significance in a study with a heterogeneous sample and/or small sample size. Clinicians scrutinizing or reporting results should place little interest on the p -value and instead consider indicators of precision, like sample size and confidence intervals. Understanding the play of chance or random sampling error is also extremely important in deciding how much weight to place on the results of a study.

13.3.1 Random Sampling Error and Sample Size

Producing a study with results that are applicable to the population is the overarching goal of clinical research. When we conduct a study, the participants are only a sample of the population. The smaller our sample, the more likely we are, just by chance, to recruit a nonrepresentative sample. The more representative your sample, the more likely that estimates of treatment effect observed with the sample will also apply to the population. As the sample size becomes larger (closer to the size of the population) or the results of repeated studies are pooled together (e.g. in a meta-analysis), the probability of a random sampling error decreases. A sample size estimate takes into consideration the variability between individuals in the population and the size of the expected difference and provides an estimate of the size of the sample required to overcome random sampling error. Although a larger sample size will help to overcome the risk of random sampling error, it cannot overcome sampling bias in the results that arise from purposefully recruiting an unrepresentative sample (e.g. selecting only those most likely to be responsive, adherent, etc.).

In the next section, we introduce the importance of including estimates of precision, like confidence intervals (CIs) when reporting or interpreting the results of a study. But even confi-

dence intervals cannot overcome random sampling error; thus, one should always be wary of the applicability of studies with a small sample size (usually defended by a sample size calculation with an unreasonably large expected effect size) or a highly selective method of sampling.

13.3.2 Confidence Intervals (CIs)

A CI represents the range in which the true score is expected to lie. Selecting a 95% CI represents means the investigator can be 95% certain that the true population mean would be contained within the interval if repeated representative samples were studied from the population. CIs should be reported no matter the summary measure being used to present the results (e.g. relative risk, odds ratio, mean difference, etc.). Then, instead of depending on the p -value to reach conclusions about the results of a study, one should interpret the lower and upper boundaries provided by the CI. If the lower and upper boundaries of the CI are implying similar conclusions, then the study may make definitive conclusions assuming the study sample is sufficiently large and representative of the population.

Fact Box 13.2: Confidence Intervals

Confidence intervals allow the reader to interpret the clinical meaningfulness of study results instead of just statistical significance. When interpreting statistical significance and CIs, consider:

If $p < 0.05$ (statistically significant), does the lower limit of the CI include differences that are not important? If it doesn't, the results are definitive in favour of treatment, while if it does, the results are uncertain.

If $p > 0.05$ (not statistically significant), does the upper limit of the CI include important differences? If it does, the results are uncertain, while if it doesn't, the treatment is definitively ineffective.

For example, the 95% CI around the relative risk of revision surgery is 3.5–8.6 (where $p < 0.05$); both sides of the CI are relaying the same message that the risk of surgery is much greater in one group compared to the other. In this situation, the study can make definitive conclusions; in a large study with a representative sample, this study convincingly demonstrates the benefit of one intervention over the other. If, on the other hand, the 95% CI around the relative risk of revision surgery is 1.01–10.5, the lower boundary of the CI is implying that there is not much risk of revision surgery in one group versus another, whereas the upper boundary of the CI is implying that there is 10× greater risk of revision surgery in one group compared to the other. In this situation, even though $p < 0.05$, the message is unclear; the study cannot make definitive conclusions.

Interpreting the CIs around the results of studies that use PROMs presents an additional challenge. Specifically, how do we know whether the difference being presented represents an important finding?

13.4 Reporting and Interpreting the Results

13.4.1 Interpreting Change in Individual Patients

To determine whether true change, as opposed to error, is the cause for different scores over time for an individual patient, we can calculate a minimally detectable change (MDC). The MDC is the threshold that defines true change over error and is calculated as the $SEM * (z\text{-score}) * \sqrt{2}$. Consider the previous example where the SEM for the hop test LSI was 4.94%. Given the results of the first hop test (LSI = 80%), you have asked your patient to undergo another 6 weeks of physiotherapy, this time concentrating on training landing biomechanics and sport-specific exercises. Upon completion of this additional rehabilitation, the new value for the LSI is 95%. For a MDC(95), the z -score is 1.96, which means that the hop test

LSI needs to change by approximately 14 cm for the clinician to be certain that the change is real and not measurement error. In our scenario, the difference in LSI between time one and time two is (95–80) 15 cm, and thus, the clinician can be certain that true change has occurred. Further, the lower boundary of the 95% CI around the second LSI is 88%, which is much closer to the desired 90% in terms of recommending that the patient returns to sport.

13.4.2 Interpreting Change Within a Group of Patients

To report the results of a within-group change, one could report the proportion of patients who have changed by an amount greater than the MDC threshold or, more meaningfully, could present the mean change from pre- to post- and use the MCID to interpret the 95% CI around the difference. For example, if the MCID for a PROM is 10 points and the average pre- to post-change score is 15 points (95%CI 5 points to 25 points), then the lower boundary of the CI implies that not all patients will achieve an important change and, thus, the study cannot conclude that the intervention will provide important improvements for all patients. On the other hand, if the 95% CI around the pre- to post-change score is 11–19 points, then (if the study is sufficiently large) the study can confidently conclude that the intervention is highly likely to provide important improvements for most patients.

13.4.3 Interpreting Change Between Groups of Patients

To report the results of a between-group comparison, one could compare the proportion of patients within each group who change by an amount greater than the MDC threshold or, more meaningfully, could present the mean between-group difference and use the MCID to interpret the 95% CI around this difference. However, the latter presents a problem because the value of the

MCID available in the literature for most PROMs will have been predominantly determined using within-group change (as described in our reliability section), which is appropriate for interpreting pre-post-intervention studies, but not appropriate for interpreting between-group studies. Here's why; consider the amount of change that will be observed from pre- to postoperative total knee replacement. Following recovery from a TKA, we expect quite remarkable improvements. For example, in a study by Kahn et al. [9] that included 172 patients who underwent TKA, the preoperative total WOMAC score was around 35 points, and the postoperative score was around 12 points; 95% CI around the mean change is approximately 20–24 points. If the MCID is approximately 15 points [4], then the study can conclude that patients undergoing a TKA will experience an important improvement.

However, what is the average difference between groups when both groups are undergoing TKA? Perhaps the comparison is between two different types of implants. In this situation, all patients are expected to benefit from TKA by an important amount. However, the research question is about measuring the difference between groups. Surely we do not expect the difference between groups to be as large as the change within groups, especially when both groups are receiving active treatment (versus no treatment). A 1993 study by Goldsmith et al. [5] determined the magnitude of the between-group MCID is significantly smaller than the within-group MCID (between 20 and 40%). Thus, if the MCID for a between-group comparison is around 5 points, then in the study by Victor et al. [24] that compared two different implants for TKA ($n = 131$) where the mean difference between groups was 3 points (95%CI 10 points to 15 points), we can determine that the study is grossly underpowered because the upper and lower boundaries of the CI include 5 (the possibility of an important difference); unfortunately, both boundaries conclude in favour of the other implant. To make a definitive conclusion that the two implants offer a similar outcome (as measured by the WOMAC), both CIs would exclude five WOMAC points.

13.4.3.1 Number Needed to Treat (NNT)

One final, and perhaps the most intuitive, method to present the results of a between-group comparison using PROMs is to provide the number needed to treat (NNT). The NNT is the average number of patients who need to be treated with the experimental treatment to prevent one additional bad outcome (e.g. the number of patients that need to be treated for one of them to benefit compared with a control in a clinical trial). It is defined as the inverse of the absolute risk reduction (1/RD). Thus, it is possible to determine the proportion of patients within each treatment group who surpass the MDC threshold and present the results as an NNT or the number of patients that must be treated with the intervention to achieve a clinically meaningful improvement in one patient (compared to the control) [12]. Thus, an NNT of five indicates that one can expect that 20% more patients will experience an improvement if they receive the experimental treatment compared to the control. The NNT is more easily interpreted because the reader does not need to understand MDC, MCID, or how to use them to interpret CIs; this work has already been done by the researcher [1].

The clinical significance of the data presented at the end of a clinical research study is dependent on the established measurement properties of the outcome measure in the sample of interest. Group differences and statistical significance are convenient ways to present data; however, investigators presenting or analysing the results should include more clinically meaningful presentations of their data to improve their readability and potential to change practice.

Take-Home Message

- The development of outcome measures has allowed clinicians to better evaluate the complex framework of health, and these instruments have become an integral part of health research; they are utilized to demonstrate patient-important changes in health and inform clinical decision-making regarding the effectiveness of treatment.

- Reliability, validity, and responsiveness are measurement properties that provide evidence for the clinician to evaluate the suitability of an outcome measure specific to the sample of interest.
- Data collected from outcome measures should be presented using statistics that are easy to interpret and relay the clinical significance of the findings.

Checklist for Interpreting Health Outcome Measures

1. What is the objective of the study?
 - (a) Was the outcome measure selected appropriate for the study objective/design?
2. Has the outcome measure been shown to be valid and reliable in the population of interest?
3. If the objective is to evaluate change, has the outcome measure been shown able to detect important change in the study population?
 - (a) Is there a reported SEM to interpret individual change?
 - (b) Is there a reported MCID to interpret group changes?
4. Are group differences presented with confidence intervals?
 - (a) Does the interpretation of the lower boundary of the 95% confidence interval offer the same conclusion as the upper boundary of the 95% CI?
 - (b) When considering the risk of random sampling error (as it relates to sample size) and sampling bias (as it relates to the representativeness of the population), what is the degree of certainty in the conclusions?
5. Has the study made appropriate conclusions and, if reasonable to do so, presented the results in terms of clinical relevance?

References

1. Bryant D, Guyatt G. Patient reported outcome measures. In: Arnold R, editor. *Pharmoeconomics*. Boca Raton: CRC Press; 2009.
2. Cochrane Collaboration glossary page [cited 2018 Jan 16]. <http://community.cochrane.org/glossary#letter-S>.

3. Deyo RA, Diehr P, Patrick DL. Reproducibility and responsiveness of health status measures. Statistics and strategies for evaluation. *Control Clin Trials*. 1991;12(4 Suppl):142S–58S.
4. Escobar A, Quintana JM, Bilbao A, Aróstegui I, Lafuente I, Vidaurreta I. Responsiveness and clinically important differences for the WOMAC and SF-36 after total knee replacement. *Osteoarthr Cartil*. 2007;15(3):273–80.
5. Goldsmith CH, Boers M, Bombardier C, Tugwell P. Criteria for clinically important changes in outcomes: development, scoring and evaluation of rheumatoid arthritis patient and trial profiles. OMERACT Committee. *J Rheumatol*. 1993;20(3):561–5.
6. Hewett TE, Myer GD, Ford KR, Heidt RS, Colosimo AJ, McLean SG, et al. Biomechanical measures of neuromuscular control and valgus loading of the knee predict anterior cruciate ligament injury risk in female athletes: a prospective study. *Am J Sports Med*. 2005;33(4):492–501.
7. Jackowski D, Guyatt G. A guide to health measurement. *Clin Orthop Relat Res*. 2003;413:80–9.
8. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407–15.
9. Kahn TL, Soheili A, Schwarzkopf R. Outcomes of total knee arthroplasty in relation to preoperative patient-reported and radiographic measures: data from the osteoarthritis initiative. *Geriatr Orthop Surg Rehabil*. 2013;4:117–26.
10. Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthritis. *Ann Rheum Dis*. 1957;16(4):494–502.
11. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Chronic Dis*. 1985;38(1):27–36.
12. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med*. 1988;318(26):1728–33.
13. Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. *Med Care*. 2000;38(9 Suppl):II84–90.
14. Marshall G, Hays R, Nicholas R. Evaluating agreement between clinical assessment methods. *Int J Methods Psychiatric Res*. 1994;4:249–57.
15. Marx RG, Menezes A, Horovitz L, Jones EC, Warren RF. A comparison of two time intervals for test-retest reliability of health status instruments. *J Clin Epidemiol*. 2003;56(8):730–5.
16. McDowell I, Newell C. *Measuring health: a guide to rating scales and questionnaires*. 2nd ed. New York: Oxford University Press; 1996.
17. Messick S. Validity. In: Linn R, editor. *Educational measurement*. Phoenix: Oryx Press; 1993. p. 13–103.
18. Paterno MV, Schmitt LC, Ford KR, Rauh MJ, Myer GD, Huang B, et al. Biomechanical measures during landing and postural stability predict second anterior cruciate ligament injury after anterior cruciate ligament reconstruction and return to sport. *Am J Sports Med*. 2010;38(10):1968–78.
19. Reid A, Birmingham TB, Stratford PW, Alcock GK, Giffin JR. Hop testing provides a reliable and valid outcome measure during rehabilitation after anterior cruciate ligament reconstruction. *Phys Ther*. 2007;87(3):337–49.
20. Rothman ML, Beltran P, Cappelleri JC, Lipscomb J, Teschendorf B, Group MFP-ROCM. Patient-reported outcomes: conceptual issues. *Value Health*. 2007;10(Suppl 2):S66–75.
21. Schünemann HJ, Guyatt GH. Commentary—good-bye M(C)ID! Hello MID, where do you come from? *Health Serv Res*. 2005;40(2):593–7.
22. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–8.
23. Streiner D, Norman G. *Health measurement scales: a practical guide to their development and use*. Oxford: Oxford University Press; 1995.
24. Victor J, Ghijssels S, Tajdar F, Van Damme G, Deprez P, Arnout N, et al. Total knee arthroplasty at 15–17 years: does implant design affect outcome? *Int Orthop*. 2014;38(2):235–41.
25. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996;34(3):220–33.
26. Wasserstein R, Lazar N. The ASA’s statement on p-values: context, process and purpose. *Am Stat*. 2016;70(2):129–33.
27. Wiebe S, Guyatt G, Weaver B, Matijevic S, Sidwell C. Comparative responsiveness of generic and specific quality-of-life instruments. *J Clin Epidemiol*. 2003;56(1):52–60.
28. World Health Organization. Preamble to the constitution of the World Health Organization as adopted by the international health conference, Geneva; 1948.



How to Document a Clinical Study and Avoid Common Mistakes in Study Conduct?

14

Caroline Mouton, Laura De Girolamo,
Daniel Theisen, and Romain Seil

14.1 Introduction

Over the last two decades, the complexity of organizing a clinical study has drastically increased. Between 1999 and 2005, the number and frequency of study procedures (measurements, questionnaires, visits, etc.) have been reported to grow at an annual rate of up to 8.7% [18]. During the same period, the number of eligibility criteria has increased by more than 12% annually [18]. As a consequence, a greater time is nowadays needed to enroll a sufficient number of

patients, and a lower retention rate of patients is observed [18]. The additional constraints to the study protocol, as well as the emergence of guidelines of best clinical practice, have thus inevitably led to a greater administrative burden and greater time duration of studies.

From an orthopedic surgeons' perspective, these changes are often perceived as a significant barrier to conduct clinical research, because this evolution is in perfect opposition to the current evolution of their clinical reality which is often dominated by an increasing economic pressure and by clinical productivity. As a consequence, the time and resources available for clinical research are decreasing. Clinical research is conducted in many centers by students or residents who often only have a limited time frame to conduct a study, although it may help them to advance in their clinical careers. Likewise, dedicated and professional staff members to support them are often not available. Nevertheless, orthopedic surgeons need to be familiar and comply with these international requirements in order to continue producing high-quality research and be able to compete with other medical specialties.

Clinical studies should be conducted in accordance with good clinical practice (GCP). A prospective study, may it be an observational cohort or a controlled trial, requires several authorizations, approvals, and/or notifications from local health authorities, Independent Ethics Committee (IEC)/Institutional Review Board (IRB), data protection authorities, insurances, and/or hospitals.

C. Mouton (✉)
Department of Orthopaedic Surgery, Centre
Hospitalier de Luxembourg—Clinique d'Eich,
Luxembourg, Luxembourg
e-mail: mouton.caroline@chl.lu

L. De Girolamo
Orthopaedic Biotechnology Laboratory, Galeazzi
Orthopaedic Institute, Milan, Italy

D. Theisen
Sports Medicine Research Laboratory, Luxembourg
Institute of Health, Luxembourg, Luxembourg

R. Seil
Department of Orthopaedic Surgery, Centre
Hospitalier de Luxembourg—Clinique d'Eich,
Luxembourg, Luxembourg

Sports Medicine Research Laboratory, Luxembourg
Institute of Health, Luxembourg, Luxembourg

Although these authorizations may differ from one country or a hospital to another, a study should not be started until all legal requirements are fulfilled.

Regulatory compliance can sometimes be laborious and time-consuming, with the consequence of delaying the start of a clinical study. However, this waiting time may be efficiently used by the investigator to establish the organizational dimension of the study. Associated with an efficient communication, it will limit time loss and allow staying on track to complete and publish the study in a timely manner. Several inadvertent consequences may occur during the study conduct in case of an insufficient preparation, an early discontinuation of the study, a delay in the recruitment of subjects, a high rate of patients lost to follow-up, low data quality, or a delay in data analysis. Greater awareness of the entire process of the study conduct and its common mistakes is therefore the key to success for a high-quality study, which may be considered for publication in a high-ranked journal.

The aim of this chapter is (1) to give an overview of the legal requirements to be respected during a study conduct as well as (2) to provide practical advice to make good use of the legally required on-site documents, to organize the study, and to avoid common mistakes in the monitoring of the study or data handling. The link between each phase of the study conduct and how it may prevent the investigator to publish will also be discussed.

14.2 On-Site Documents: Definitions and Use

On-site documents should demonstrate that the principal investigator (PI) is conducting the study according to standards and applicable regulatory requirements. In case of an audit, they can be reviewed by the legal authorities or the sponsor(s). The minimum list of essential documents to be available on-site during the study conduct is reported here [1]. This list mainly includes documents which have been previously submitted to authorities. Guidance on how to properly use these documents for the study conduct is provided in this chapter [2].

14.2.1 Protocol, Case Report Form (CRF), Information to the Patient, Informed Consent Form (ICF), and Amendments

This section should include the initial study protocol as submitted to the IEC/IRB as well as all available subsequent amendments. The latter are documented changes to the protocol initially approved by the IEC/IRB.

The CRF is a printed or electronic document used to record all data required by the study protocol such as patient demographics data, study interventions, outcomes, and adverse events. It should not be confused with source documents which will be described later in this chapter. A blank copy of the case report form (CRF) as well as a blank copy of patient-related outcomes measures (PROMs) and adverse event (AE) report must be available on-site. This section of the binder should also include the latest version of all documents used for the recruitment of participants, that is to say the information provided to the patient, the informed consent form (ICF), and all advertisements used to recruit participants (newspaper, radio, posters, and flyers) [16].

As for the protocol, the CRF or the information to the patients may be subjected to changes during the study conduct (i.e., following an amendment in the protocol). To avoid time loss, it is important to evaluate properly how a possible change in the study protocol may affect the other study documents (i.e., CRF, PROM, and AE) and submit all new versions together.

14.2.2 Approval/Favorable Opinion of Independent Ethics Committee (IEC)/Institutional Review Board (IRB) and Annual Reports

On-site binders must include study interim or annual reports as well as all documents received from the IEC/IRB to certify that the study protocol and its amendments have been reviewed and approved by legal authorities.

14.2.3 Signed Informed Consent Forms (ICF)

Participation in a study must be on a voluntary basis; the patient's decision must not be influenced by the investigator/staff. If the patient is willing to participate into a study, the informed consent is documented by means of a written, signed, and dated ICF. The latter attests that the study was explained to the patient and understood and that his/her consent was freely given. When a subject is not capable of giving informed consent, the permission of a legally authorized representative should be obtained in accordance with the applicable law (i.e., for minors, the parents must sign on his/her behalf).

14.2.4 Source Documents

Source documents are original documents, data, and records (i.e., hospital/medical records, laboratory notes, X-rays, etc.). They should be attributable, legible, contemporaneous, original, and accurate (ALCOA) [6] and distinguished from the CRF. For each data collected, the source documentation (medical file) should be explicitly mentioned to enable the PI or an independent observer to reconfirm the data (even years after completion of the study).

14.2.5 Signed, Dated, and Completed CRF

If study data are recorded on paper during study visits, the documentation must be retained and documented in this section. For electronic CRF (eCRF), results can be entered directly into the eCRF by an authorized staff member to avoid paper transcription. In this case, the eCRF is the source. Guidance has been made available for electronic source data by the FDA to clarify when the CRF is considered as the source document or not [4].

14.2.6 Adverse Events (AE) and Notification of Serious Adverse Events (SAEs)

An adverse event (AE) is “any untoward medical occurrence, unintended disease or injury, or untoward clinical signs (including abnormal laboratory findings) in subjects, users or other persons, whether or not related to the investigational medical device” [1]. All AE should be reported during the study conduct. If AE are integrated into an eCRF, the same rules than for the CRF apply. Furthermore, serious adverse events (SAEs) must be reported to the IRB/IEC and to the hospital within a given time frame from the occurrence of the adverse event, as well as to sponsors if applicable. They include any event related or not to the participation in the study (i.e., related or not to the study device/technique/drug) that has either led to the death of the patient or that has resulted in a life-threatening illness or injury, in a permanent impairment of a body structure or function, or in a hospital admission. Any medical intervention performed to prevent illness, injury, or impairment is also considered to be a SAE.

Clinical Vignette 1: Discussion on Adverse Events

Ms. X comes for a medical visit 3 months after total knee replacement. She also has consented before the surgery to participate in a post-marketing study. At the physical examination, the doctor detects knee stiffness that requires manipulation under anesthesia. The clinical coordinator is present, and the doctor asks for the adverse event (AE) form. After discussion, both the doctor and coordinator agree to report a serious adverse event (SAE). Although knee stiffness is a known complication after total knee arthroplasty, it led to the patient hospitalization, and this intervention is useful to prevent further impairment. The SAE is reported both to the sponsor and the hospital although the event may not be directly related to the medical device (knee arthroplasty).

14.2.7 Subject Screening, Enrollment, and Dropout Log

The screening and enrollment logs report all patients screened and informed about the study. It is used to report the outcome of the screening (as screening failures can occur). It usually includes the following items: investigator name, site, patient initials, date informed consent signed, version of consent, date of screening, reasons for screening failure, reason for withdrawal/exclusion, study code assigned, and staff initials. Any withdrawal or lost to follow-up during the study conduct will be reported separately on the dropout log.

Logs can help to conduct the study in a more efficient and organized manner. Documentation of the reasons why the screening has failed (i.e., patient refused to participate in the study, patient older than the upper limit of age considered) can provide information regarding the ability of the study team to enroll patients and retain them. This information may be useful to adapt the recruitment period.

14.2.8 Subject Identification Code List

This log enables the PI to track the correspondence between enrolled subject name and allocated study code. This list is kept confidential, well secured, and/or protected by password if it is digital. Only the PI and delegated staff should have access to this list.

14.2.9 Study Staff and Training Log

The PI can delegate tasks to his staff through the signature sheet. The latter documents signatures and initials of all persons authorized (co-investigators) to screen subjects, assess inclusion/exclusion criteria, gather patient consent, make entries and/or corrections on CRFs, perform measurements, etc. It can be completed with a training log to prove that all staff members have been trained for the study procedures.

14.2.10 Curriculum Vitae of Principal and Co-investigators

These documents are helpful to document qualifications and eligibility of investigators to conduct a study and/or provide medical supervision of subjects.

14.2.11 Other On-Site Documents According to the Study

In addition to the previously cited documents, if the protocol includes known technical procedures, it is recommended to report any update that may occur during the study conduct in testing procedures, including updates on normal value(s), certification, accreditation, or quality control. From a scientific point of view, it may help to trace any outlier/abnormal value when analyzing the data and help to correctly interpret the data.

If biological samples are collected and stored during the study, it is important to document the location and identification process of retained samples in case a test has to be repeated.

If the study includes the investigation of a product (i.e., medical device), the investigator's brochure and updates, instructions for handling investigational product, marketing authorization (i.e., CE mark), and procedures for shipment and storage must be available on-site.

Finally, if the study is being sponsored, other documents may apply, such as an up-to-date insurance, financial agreements between involved parties, monitoring visit reports to document visits, and findings of the monitor and any relevant communication other than site visits including letters, meeting notes, notes of telephone calls, etc.

14.3 Monitoring the Study

Once the regulatory binders contain all required documents, the study can theoretically start. Beyond GCP requirements, it may be recommended to conduct the study according to the

Fact Box 14.1

Clinical studies should be conducted in accordance with good clinical practice (GCP). Any substantial change to the protocol during the study conduct requires approval from the Independent Ethics Committee (IEC)/Institutional Review Board (IRB).

Consent is documented by means of a written, signed, and dated informed consent form (ICF) which attests that the information was explained and understood and that consent was freely given.

Source documents are original documents, data, and records from the hospital, while the case report form (CRF) is a printed or electronic document used to report pseudonymized data according to the study protocol (i.e., the name of the patient does not appear on the document, only the study code allocated to the patient for his/her participation). A list to track the correspondence between enrolled subject name and allocated study code must be kept by the PI separately.

Subject screening, enrollment, and dropout logs must be used to report screening of patients, outcomes of screening, and any withdrawal or lost to follow-up.

All adverse events (AE) should be reported (all abnormal findings during the patient follow-up). Serious adverse events (SAEs) must be reported to the IRB/IEC and to the hospital within a given time frame (usually within the first 24 h) from the occurrence of the adverse event.

The PI can delegate tasks to his staff, which should be documented by a signature sheet and training log.

quality of manuscript foreseen. For example, the PI must be aware that a loss to follow-up greater than 20% will inevitably lead to a lower level of evidence of the results. He/she should thus ensure that the follow-up of patients as well as all aspects

considered during the review process of a manuscript is correctly addressed. Existing checklists and statements to report studies may therefore be consulted before the study starts to properly plan the latter [9, 10, 25].

14.3.1 Responsibilities of the PI

The PI assumes the responsibility for proper conduct of the study. He/she is responsible for protecting the rights, safety, and welfare of subjects under his/her care during a clinical study. During the study conduct, whether tasks are delegated or not, the PI should ensure that:

- The conduct of the study is in compliance with the protocol.
- No deviation is allowed from the protocol without prior favorable opinion from the IRB/IEC (protocol amendment).
- Any deviation from the protocol is documented and explained (i.e., a deviation may occur and be unavoidable if the PI wants to eliminate immediate hazard(s) to the participant).
- Medical devices or drugs, if investigated, are used in accordance with the approved protocol.
- Codes in randomized controlled trials are broken only in accordance with the protocol (unblinding).
- Obtaining and documenting patient consent are performed according to the applicable regulatory requirement.
- Data in the CRFs and in all required reports are accurate, complete, and legible.
- A summary of the study status is reported annually to the IRB/IEC.
- All SAEs are reported immediately to the regulatory authorities.

The above list is not exhaustive but includes the main responsibilities of the PI during study conduct. Further guidance on investigator responsibilities has been made available by the FDA [5].

14.3.2 Communication and Delegation

Communicating with the research team (especially when delegating) before and during the study conduct is critical. On-field employees can be a tremendous source of feedback (protocol deviations, difficulties, experiences, and adverse events) and ideas. Ideally, the PI should first inform all persons involved in patient care (nurses, physios, radiologists, secretaries, etc.) that a study is ongoing. It can be achieved by transmitting the synopsis of the study and/or by organizing an introduction session.

The PI may also consider informing the general practitioner (GP) of the participant through a doctor-to-doctor referral letter. This is a common practice to inform the GP about the treatment his/her patient underwent and thus to detect any possible adverse events at an early stage.

During a study, dedicated research staff may be required to support clinical staff and patients. To comply with all activities induced by the clinical study, the PI can delegate tasks provided that the study staff log explicitly mentions it and that delegates are trained. The delegates should have full knowledge of the protocol and should be familiar with the condition studied. The PI should furthermore encourage his staff to take advantage of the numerous online GCP training courses to obtain the GCP certificate. The PI may also plan training sessions for new procedures.

14.3.3 Enrollment of Subjects into the Study: Screening, Recruitment, Eligibility, and Informed Consent

Informed consent is the process by which a subject voluntarily confirms his/her willingness to participate in a study. It must be obtained from each participant prior performing any specific study procedures. The latter include any activity not included in the standard treatment.

To plan enrollment, several questions may be asked:

14.3.3.1 How Will Eligible Patients Be Identified?

The prescreening phase helps to identify potentially eligible patients. If the study includes a surgical procedure (i.e., anterior cruciate ligament reconstruction or total knee replacement), given that time between indication and surgery is sufficient for the patient to consider his/her participation, the easiest way may be to identify patients through the surgical planning. If only the PI can identify patients with the pathology of interest, only he/she may provide the information to the patient or inform the study staff team to enroll the patient. In any case, inclusion/exclusion criteria should be applied strictly.

14.3.3.2 What May Limit the Recruitment?

Enrollment during the first month is a strong predictor of study completion [20]. It is thus critical to correctly plan how eligible patients will be contacted and recruited. From the clinician's point of view, barriers to recruitment may include underestimation of the prevalence of the condition studied requiring longer enrollment time to reach the desired sample size [21], time constraints, lack of staff and training, and difficulties with the consent procedure [26]. From the patient's perspective, additional demands of the study compared with the routine care and concerns about information, data privacy, and consent may constitute some factors that limit their willingness to participate in a study.

14.3.3.3 How Will Eligible Patients Be Contacted?

For nonintervention studies (i.e., observational studies), an "opt-out" recruitment strategy (contact the patient instead of waiting that the patient expresses his/her willingness to participate) is advised [22]. Compared to emails and letters, direct phone calls by the investigator(s) are recognized as the most effective method for recruitment [27].

A letter signed by the PI together with the information may also previously be sent to the patient (or given during the medical visit by the PI). The letter could indicate to contact the dedi-

cated person of the team for further question on the study and/or inform that he/she will be contacted by a team member (naming a specific person may here be useful so that the patient is waiting for this phone call).

14.3.3.4 How to Report Prescreened Patients?

Screening logs should remain restricted in content but should document as many eligibility criteria as possible [15]. Since there is no informed consent from the patient, research procedures or interventions should not yet take place. Practically, the site may create a prescreening sheet to follow track for screened/contacted patients. This sheet should include screening date, eligibility criteria that can be assessed using the medical record (i.e., patients planned for ACL reconstruction, age, type of graft planned), and contact date. Only procedures that are performed as part of the routine clinical practice may be looked at, and only results used for determining study eligibility should be screened before obtaining consent [17].

14.3.3.5 How to Instruct the Patient About the Study?

Participants should receive information both in written and oral formats (call, visit) in a nontechnical language. The PI or the delegated person should capture the patient's perspective and take time to answer his/her questions. Patients are indeed less likely to enter studies that they find difficult to understand and that require multiple follow-ups [29]. It may be worth mentioning whether visits are part of the clinical routine or not and explicitly mention additional visits and procedures. Patients may indeed not realize that some procedures will solely be performed for research purposes and are not required for their medical care.

Information given to patients should include information about:

- The study: its purpose, participation duration and number of visits, subject's involvement and responsibilities, side effects, risks and benefits, and alternatives to treatment.

- The participation to the study: participation on a voluntary basis, the possibility to refuse to participate and to withdraw from the study at any time without penalty or loss of benefits, any compensation and amount if applicable, reasons for possible early termination of participation/study (i.e., patient not compliant with visits, medical condition interfering with study protocol), and new information if new findings become available that may be relevant to the subject's willingness to continue participation in the study.
- Data protection: confidentiality of personal information recorded and records identifying the subject, protection of privacy, and access to pseudonymized data from third parties if applicable.
- Additional medication administered if needed, procedure, and insurance if a damage occurs.
- Approval of the study by EC/IRB.
- Principal investigator/funder of study/contact person for any question related to the study and in case of adverse event

14.3.3.6 How to Record Patient Consent?

Informed consent is documented by means of a written, signed, and dated ICF. Each consent form (usually a minimum of 2) must be dated and signed both by the patient and the investigator. The PI will be responsible for any misconduct on the process of ICF (wrong date, falsified signature, missing consent). One consent is then kept for the regulatory binders, and one is given to the participant. The hospital may also require a copy for the electronic health record.

14.3.3.7 What If New Information About the Medical Device/ Drug Becomes Available During the Study or If the Protocol Procedure Changes?

If new information on risk and/or benefits arises or if major protocol amendments occur during the study, the PI should ensure that subjects are informed and re-consent to participate in the study.

14.3.3.8 What to Do If the Recruitment Is Slower as Expected?

Poor subject recruitment and retention is 1 of the 15 common reasons for failure in clinical research [11]. The PI should be able to predict recruitment rate according to his facilities, patients, and protocol. He/she should also be able to identify any factors that could prevent the team to properly recruit for the study. During the study conduct, the PI should thus review the recruitment on a regular basis (i.e., with the help of the screening log). If the recruitment rate is lower than expected, a discussion with the study staff may help to identify the difficulties met in practice. The study team may also consider to adapt the protocol and information given to the patient if too complex or leading to confusion (i.e., too many visits, divergence from routine care) or to increase the duration of the recruitment period [29].

14.3.4 Study Visits: Compliance with the Approved Protocol and Protocol Amendments

14.3.4.1 Study Visits

The PI should consider providing the study staff with a brief worksheet/checklist indicating procedures to be performed at each visit. It can serve as a reminder and ensure that procedures are completed in a timely manner without any missing data.

To efficiently organize the study visits, the study team should consider the following aspects: visit windows (range of days in which a subject visit can occur according to the study protocol), room and staff availability, availability of support department (i.e., X-ray), and whether procedures are part of the clinical routine or not.

The study staff should be careful to avoid long waiting times for study visits. Ideally, the visits should coincide with routine visits if applicable and be short [23]. The follow-up visits may also be proposed at a location convenient for the patient and/or outside working hours. The appointments should be scheduled as soon as possible (i.e., at discharge if any surgery), and the

team should keep track of visits and windows in a sheet/file. Questionnaires to be filled in by the patient, if applicable, may be sent some time before the visit, together with a reminder of the appointment. If the patient does not come to the visit, the study staff should make all efforts possible to contact the patient to avoid a loss to follow-up. These efforts should be documented. Any definitive loss to follow-up should be reported in the dropout log.

14.3.4.2 Compliance with Protocol and Protocol Amendments

Adherence to the study protocol is essential. Any deviation from or violation of the protocol should be documented [8, 28]. Deviations from the protocol are defined as changes or noncompliance with the study protocol that does not have a significant effect on the participant's rights, safety, or welfare (i.e., missing visits or data). Protocol violation may affect the participant's rights, safety, or welfare (i.e., inclusion/exclusion criteria not met, failure to obtain valid informed consent).

If substantial protocol modifications become necessary during the study conduct (either to avoid protocol deviation/violation or for other reasons), amendments to the protocol must first be approved by the IEC/IRB. Overall, about two-thirds of protocols are reported to require one or more amendments [14, 19] although the latter have an additional impact on study costs, timeline, and resources [19, 24]. One-third of these amendments are considered to be avoidable if inconsistencies/errors in the protocol and difficulties in recruiting study volunteers are better anticipated.

The definition of "substantial" may vary according to legal authorities but generally includes all modifications that may impact the safety or physical or mental integrity of the subjects or the scientific value of the study. According to the EU guidance (Sects. 14.3.3 and 14.3.4) [12], substantial amendments include (non-exhaustive list):

- Amendments to the protocol: Any changes in the population studied, procedures, and moni-

toring visits including changes of the main objective or endpoints or changes in the recruitment procedure (inclusion/exclusion criteria, additional group of patients, etc.) are considered as substantial changes. This does not include modification of the title, addition/deletion to tertiary endpoints, minor increase in the duration of the study (<10% of the overall time), or increase of >10% of the overall time of the study provided that monitoring visits are unchanged (i.e., no additional visit).

- Amendments concerning the product studied: any new information on the study product or any new information made available by the manufacturer.
- Amendments to other documents: any changes of sponsor/principal investigator or revocation of the product marketing authorization.

Substantial amendments must be approved by the IEC/IRB. As for non-substantial amendments, it is possible to record them and submit them simultaneously with the notification of a substantial amendment or at least inform the IEC/IRB.

For each new protocol version, date and version identifier should be properly reported, and modifications should be highlighted, and a detailed summary of protocol changes including old text, new text, and rationale for change should be provided (SPIRIT guidelines—Item 3—date and version identifier) [10]. It is usually recommended to maintain both a track change and a clean version of the protocol to the IRB/IEC including a document history.

14.3.5 Safety Management and Reporting

The PI should ensure that the risk-benefit for the patient to participate in the study remains constantly favorable. He should ensure the communication with the study staff who is often the first to observe unanticipated risks.

All AE should be followed, and detailed written reports should be provided until the risk is eliminated or AE resolved. If judged necessary by the authorities, the protocol may be amended and ICF updated to inform patients about new risks. Alternatively, the study may be terminated prematurely.

Fact Box 14.2

The PI assumes the responsibility for proper conduct of the study even if tasks are delegated.

Communicate not only with the study team, but inform all departments/doctors involved in the patient's care about the ongoing study.

Aspect to organize for an efficient study conduct:

- Identification of eligible patients.
- Contact with patients and information about the study (written and oral).
- Consent signature and on-site documentation (screening log, study code allocated to the patient).
- Study visits (according to visit windows, room and staff availabilities, and clinical routine).
- Procedures to be performed at each visit.
- Contact with patients missing a visit.

14.4 Managing Study Data

Data management includes all procedures for collecting, handling, manipulating, analyzing, and storing/archiving data used during the study conduct. All information should be recorded, handled, and stored in a way that allows its accurate reporting, interpretation, and verification.

14.4.1 Data Quality and Integrity

Quality and integrity are related. If data quality is bad, data integrity cannot be reached. Data quality refers to the essential characteristics of data. These should be attributable, legible, contemporaneous, original, and accurate (ALCOA) [6]. Data integrity refers to the validity and consistency of data. Mechanisms should be in place to prevent accidental modification or erasure of the data (i.e., data backup).

The design of the CRF (paper or electronic form) is a key quality step in ensuring the data required by the protocol [7]. To ensure data quality and integrity, the PI should ensure that the CRF is standardized (i.e., format of date, pick lists).

Care should also be given to provide the study staff on how to deal with:

- Confidentiality: The CRF should not include any information that can be used to identify the study participant.
- Missing data: No field should be left blank. Visits that the participants fail to make, tests not conducted, and examinations not performed or missing information should be reported by indicating “not done,” “not applicable,” or “unknown.”
- Completion, change, or correction on a CRF: In any format (paper or electronic), it is required to record who entered and generated the data. Any change or correction should be dated, initialed, and explained. Original data should be crossed out with a single line that leaves the original information clearly visible. The correct data should be inserted next to the erroneous data, and the form should be initialed and dated. Similar systems should be activated for electronic forms.

Identification of missing data and/or discrepancies should be performed on a regular basis during the study conduct, especially if CRF completion is delegated to staff members. Regularly, the PI should also control the data-

base for inclusion/exclusion criteria, validity of data, and outliers. At the end of the patient follow-up, CRF should be completed and signed. At the end of the study, records related to the study should be kept for the period of time required by national/local laws and regulations.

14.4.2 Data Access, Confidentiality, and Privacy

Since May 2018, the General Data Protection Regulation (GDPR) replaces the Data Protection Directive 95/46/EC to protect data privacy from European residents [3]. The implication of such regulation on research is developed in another chapter [13], and only critical aspects for study conduct will be recalled in this section.

Privacy implies that the participant information will be protected and not disclosed without the knowledge/permission of the participant himself. Confidentiality involves that the PI and his team protect the information on the patient from deliberate or accidental disclosure and follow procedures to release the information only to authorized parties.

The PI must protect the confidentiality of information retrieved from medical records and visits. For example, he/she should consider encrypting data, restricting access to study records, keeping study records in secured areas, and maintaining subjects' names and study code separately (subject identification code list).

The PI or the delegate should inform the study participant in written (ICF) and oral format about who will have access to his/her personal data (study team, IRB/IEC, regulatory authorities, sponsors if applicable) and about the measures taken to ensure the confidentiality and security of personal information. The participant should also be aware that his personal data will be kept confidential and will not be publicly available even if the results are foreseen to be published in a scientific manuscript.

Fact Box 14.3

- Respect the confidentiality of data on the CRF: Use the participant study code.
- Control data: Identify regularly missing data, control inclusion/exclusion criteria, and check for abnormal values. All information should be accurate and verifiable (data entered in the CRF should be identical to the data in the patient medical record). No field should be left blank (report “not done,” “not applicable,” “unknown”). Any change or correction of the CRF should be dated, initialed, and explained.
- Respect privacy: Only release information to authorized team and parties (i.e., sponsor if applicable).

Take-Home Message

- Principal investigators are responsible for the entire study conduct, independent of whether study related tasks are delegated or not.
- Compliance with legal requirements and GCPs should be ensured during the entire duration of the study.
- Regulatory binders should be up to date, the informed consent process should be performed according to the participants’ rights, and data confidentiality and privacy should be maintained.
- The organization of the study conduct from patient screening to data quality will inherently follow the Deming circle: Plan, communication with the team, agreement on task delegation, planning of patient enrollment and study visits, and respect of confidentiality; Do, recruitment, data collection, and restriction of the access to study records; Check, regular discussion with the team to identify problems (i.e., adverse events, low rate of enrollment) and regular check of the database to correct for missing data and inconsistencies; and Act, adapt procedures if necessary.

Bibliography

1. ICH tripartite guideline for good clinical practices E6 (R1). 1996. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R1_Guideline.pdf.
2. NCCIH clinical research toolbox—essential documents/regulatory binder. <https://nccih.nih.gov/grants/toolbox#binder>.
3. (2016) regulation (EU) 2016/679 of the European Parliament and of the council of 27 April 2016 on the protection of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal L 119, 4.5.2016, p. 1–88. Official Journal L 119 1–88.
4. Administration USDoHaHSFaD. Guidance for industry—electronic source data in clinical investigations. 2013. <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM328691.pdf>.
5. Administration USDoHaHSFaD. Guidance for industry—investigator responsibilities—protecting the rights, safety, and welfare of study subjects. 2009. <https://www.fda.gov/downloads/Drugs/.../Guidances/UCM187772.pdf>.
6. Bargaje C. Good documentation practice in clinical research. *Perspect Clin Res.* 2011;2(2):59–63.
7. Bellary S, Krishnankutty B, Latha MS. Basics of case report form designing in clinical research. *Perspect Clin Res.* 2014;5(4):159–66.
8. Bhatt A. Protocol deviation and violation. *Perspect Clin Res.* 2012;3(3):117.
9. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig L, et al. STARD 2015: an updated list of essential items for reporting diagnostic accuracy studies. *BMJ.* 2015;351:h5527.
10. Chan AW, Tetzlaff JM, Gotzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ.* 2013;346:e7586.
11. Clark GT, Mulligan R. Fifteen common mistakes encountered in clinical research. *J Prosthodont Res.* 2011;55(1):1–6.
12. Communication from the commission—detailed guidance on the request to the competent authorities for authorisation of a clinical trial on a medicinal product for human use tnos.
13. Theisen D, Moksnes H, Hardy C, Engebretsen L, Seil R. How to organise an international register in compliance with the European GDPR: walking in the footsteps of PAMI (Paediatric ACL Monitoring Initiative).
14. Decullier E, Lhéritier V, Chapuis F. The activity of French Research Ethics Committees and characteristics of biomedical research protocols involving humans: a retrospective cohort study. *BMC Med Ethics.* 2005;6:9.

15. Elm JJ, Palesch Y, Easton JD, Lindblad A, Barsan W, Silbergleit R, et al. Screen failure data in clinical trials: are screening logs worth it? *Clin Trials*. 2014;11(4):467–72.
16. FDA. Guidance for Institutional Review Boards and clinical investigators—recruiting study subjects—information sheet. 2018. <https://www.fda.gov/RegulatoryInformation/Guidances/ucm126428.htm>.
17. FDA. Guidance for Institutional Review Boards and clinical investigators—screening tests prior to study enrollment—information sheet. 2018. <https://www.fda.gov/RegulatoryInformation/Guidances/ucm126430.htm>.
18. Getz KA, Wenger J, Campo RA, Seguire ES, Kaitin KI. Assessing the impact of protocol design changes on clinical trial performance. *Am J Ther*. 2008;15(5):450–7.
19. Getz KA, Zuckerman R, Cropp AB, Hindle AL, Krauss R, Kaitin KI. Measuring the incidence, causes, and repercussions of protocol amendments. *Drug Inf J*. 2011;45(3):265–75.
20. Haidich AB, Ioannidis JP. Effect of early patient enrollment on the time to completion and publication of randomized controlled trials. *Am J Epidemiol*. 2001;154(9):873–80.
21. Hewison J, Haines A. Overcoming barriers to recruitment in health research. *BMJ*. 2006;333(7562):300–2.
22. Junghans C, Feder G, Hemingway H, Timmis A, Jones M. Recruiting patients to medical research: double blind randomised trial of “opt-in” versus “opt-out” strategies. *BMJ*. 2005;331(7522):940.
23. Kaur M, Sprague S, Ignacy T, Thoma A, Bhandari M, Farrokhyar F. How to optimize participant retention and complete follow-up in surgical research. *Can J Surg*. 2014;57(6):420–7.
24. Lösch C, Neuhäuser M. The statistical analysis of a clinical trial when a protocol amendment changed the inclusion criteria. *BMC Med Res Methodol*. 2008;8:16.
25. Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c869.
26. Ross S, Grant A, Counsell C, Gillespie W, Russell I, Prescott R. Barriers to participation in randomised controlled trials: a systematic review. *J Clin Epidemiol*. 1999;52(12):1143–56.
27. Schroy PC 3rd, Glick JT, Robinson P, Lydotes MA, Heeren TC, Prout M, et al. A cost-effectiveness analysis of subject recruitment strategies in the HIPAA era: results from a colorectal cancer screening adherence trial. *Clin Trials*. 2009;6(6):597–609.
28. Sweetman EA, Doig GS. Failure to report protocol violations in clinical trials: a threat to internal validity? *Trials*. 2011;12:214.
29. Thoma A, Farrokhyar F, McKnight L, Bhandari M. Practical tips for surgical research: how to optimize patient recruitment. *Can J Surg*. 2010;53(3):205–10.



Framework for Selecting Clinical Outcomes for Clinical Trials

15

Adam J. Popchak, Andrew D. Lynch,
and James J. Irrgang

15.1 Outcomes in Healthcare

Outcomes in healthcare exist in various domains, such as clinical outcomes, process-of-care outcomes, patient satisfaction outcomes, and cost outcomes. Therefore, when selecting outcomes for clinical trials, one must determine which outcome is meaningful in the context of the trial. Clinical outcomes may relate to impairments in body structure and function, activity limitations, or participation restrictions. Process-of-care outcomes are often related to utilization of resources, duration of care, and the procedures and interventions provided. Patient satisfaction may be related to satisfaction with the healthcare provider, the support staff, or the result of care. Finally, cost outcomes often focus on the direct costs of the medical care or the indirect costs of the illness or pathology. In regard to cost, payment reform in healthcare is a driving factor of what a meaningful outcome is, with value-based payments replacing volume-based payments. Value, in healthcare, is the outcomes that are achieved compared to the costs to achieve them [29]. Increased value is associated with the improvement of the health outcomes achieved and the consideration of costs. Therefore, to mea-

sure value, there is an inherent requirement to accurately measure health outcomes.

Patient-centred outcomes measure the result of medical care from the perspective of the patient. Patient-reported outcomes (PROs) commonly measure the patient's perception of their symptoms, activities, or participation levels. Therefore, when selecting outcomes for clinical trials, determination of what is important to the population of interest is essential. Relationships between impairments [26, 36] and resulting activity limitations and participation restrictions that affect patient-centred outcomes are not always direct and vary amongst individuals. Additionally, activity and participation are of utmost concern to the individual. Therefore, measures of activity and participation should be the primary outcome measure in clinical trials concerned with the outcome of patient care.

Practical considerations for choosing an outcome include the purpose of the measurement; the relevance to the patient population; the psychometric properties such as reliability, validity, and responsiveness; and the clinician or respondent burden. Additional considerations include if the purpose of the measurement is to discriminate between subjects or groups, predict current or future status, or evaluate change in a condition over time [18]. Ideally, the measurement matches the level of intervention in a given clinical trial. Interventions aimed at treating an impairment should have outcome measurements that can

A. J. Popchak (✉) · A. D. Lynch · J. J. Irrgang
Departments of Physical Therapy and Orthopaedic
Surgery, University of Pittsburgh,
Pittsburgh, PA, USA
e-mail: ajp64@pitt.edu

evaluate outcomes at the level of an impairment. Likewise, a trial with a purpose of reducing a disability should have an outcome measure capable of assessing levels of disability. Ultimately, the outcome measure should serve the purpose of the research trial to provide information necessary to draw a conclusion (Fig. 15.1).

15.2 Measures of Activity and Participation

There are two main approaches to assess activity and participation, performance-based testing and patient-reported measures [21]. Performance-based measures (PBM) rely on a rater’s assessment of a patient’s performance on specific physical tasks [21]. Performance may be assessed qualitatively with rating scales for the level of assistance needed or quality of movement or quantitatively based on the amount of time required to complete a task, the tolerance for completing a task, or the measurable results of a test such as jumping distance, balance, speed, or time. For example, when considering outcomes related to movement in the context of acute care or rehabilitation settings, level of assistance works particularly well. Advantages of PBM of physical function compared to patient-reported measures include superior reproducibility, greater sensitivity to change (responsiveness), and less vulnerability to external influences or biases [8, 21].

Patient-reported outcomes (PROs) are completed by the patient or a proxy (e.g. a parent for

a child) that rely on self-perception of the symptoms, impairments, and abilities [21]. Patient-reported outcomes of health-related quality of life (HRQOL) can be either general or specific, with pros and cons being present with each type of measures.

Generic health status measures are applicable to diverse populations and usually measure multiple aspects of health such as physical, emotional, and social. The common examples of a general health measure are the Medical Outcomes Study 36-Item Short Form (SF-36) [25] and 12-Item Short Form (SF-12) [35]. A more contemporary measure is the PROMIS Global-10 [9], which utilizes item response theory (see chapter on adaptive testing). General health measures permit comparisons across populations with different health conditions as well as being more likely to detect unexpected effects of an intervention. However, they are less responsive than specific measures of health status; are susceptible to being unable to distinguish between groups secondary to their scores being either as good as possible or as bad as possible in both the treatment and control groups when used for high or low functioning individuals, otherwise known as ceiling and floor effects; generally have content that is less relevant to the patient and clinician; and tend to be longer and more difficult to score.

Specific health status measures focus on content specific to the primary condition or population of interest, potentially creating a more responsive instrument. To achieve this, specific health status measures include only the aspects of

Fig. 15.1 Determining the primary outcome measure

Level of Intervention	Level of Outcome Measurement		
	Impairment	Activity Limitation	Participation Restriction
Impairment	✓	X	X
Activity Limitation	X	✓	X
Participation Restriction	X	X	✓

HRQOL that are relevant to the condition or population being studied. Specific health status measures have the distinct advantages of improved responsiveness, lower respondent burden, are easy to score and interpret, and are more likely to be accepted by patients and clinicians secondary to having greater relevance to the condition of interest. However, specific health measures do not measure all aspects of health that may influence the overall status, nor do they allow for comparison between different disease states and/or populations.

Disease-, region-, or patient-specific scales exist as specific health status measures. Disease-specific scales are designed for a particular disease process or pathology. The content reflects symptoms, activity limitations, and participation restrictions that are experienced by the individual with the disease. Examples of disease-specific scales used include the Lysholm [33] and Cincinnati Knee Rating Scale [27] (knee ligament scales), the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC) [4] (osteoarthritis-specific scales), and the Western Ontario Rotator Cuff Index (WORC) [17] (rotator cuff dysfunction).

Region-specific scales are designed for use on a wide variety of disorders or impairments that affect a particular region. The content reflects all possible symptoms, activity limitations, and participation restrictions that can arise from impairment of a specific region. Examples of region-specific scales include the Neck disability Index (NDI) [34]; Penn Shoulder Score (PSS) [22]; Disabilities of the Arm, Shoulder, and Hand (DASH) [12]; Oswestry Disability Index (ODI) [7]; and the Knee Outcome Survey Activities of Daily Living Scale (KOS-ADLS) [15].

Patient-specific scales are defined by the patient, often with the patient providing a list of 3–5 relevant activities they are unable to do or are have difficulty performing. Generally, the activities are then given a numerical value rating, often on an 11-point scale, where 0 is “unable to do” and 10 is “able to do at preinjury level”. Specifically, the Patient-Specific Functional Scale (PSFS) is used primarily in patients with musculoskeletal disorders and can be used in

patients with varying levels of independence [11]. The PSFS has shown to be a valid, reliable, and responsive outcome measures for a variety of musculoskeletal problems [10]. Patient-specific scales are applicable to a large number of conditions, are efficient and easy to administer, and have adequate psychometric characteristics, in particular responsiveness to change over time. However, patient-specific scales limit comparisons between patients secondary to the lack of uniformity of content items that are determined by each individual.

As outcome measures are being utilized in clinical trials, the measures should be standardized to ensure the ability to compare results, thereby working with a “common currency” effect. In the absence of such standardization, results and conclusions of studies are unable to be compared, and additional evidence to support or refute a conclusion remains missing. There has been some effort to document core outcome sets for a body region or condition. A core outcome set establishes a minimum data set that should be collected for a particular condition to allow for comparison amongst studies.

15.3 Psychometric Considerations

Psychometric considerations are vitally important when selecting outcome measures. Primarily, investigators must be concerned with the outcome measure’s reliability, validity, and responsiveness. Additional considerations include the purpose of the measurement, the relevance to the patient population, and the clinician or respondent burden. All of these factors should be considered when selecting an appropriate outcome measure for clinical trials.

15.3.1 Reliability

Reliability is the consistency of measurement and how much error one can expect in the chosen outcome. Acceptable reliability levels are necessary to ensure that the error associated with the measurement is small enough to detect

actual changes in what is being measured [31]. Therefore, reliability can be conceptualized as the dependability or the predictability of a measure [30]. Reliability is fundamental to clinical research. In the absence of it, the researcher cannot be confident in the data that is collected, and no definitive conclusion can be made from it [30].

Reliability can be assessed on multiple levels. When assessing whether multiple items measure the same construct, such as with questionnaires and interviews, reliability is related to internal consistency [30]. Internal consistency is the degree to which all items on a scale consistently measure the underlying condition [30]. Internal consistency applies to measures that consist of multiple items and therefore is related to errors of measurement that are linked to content sampling. Two primary approaches exist to measure internal consistency, split-half reliability, and test item reliability [30]. Split-half measures the extent to which all parts of the test contribute equally to what is being measured and is based on dividing the items on an instrument into two halves and correlating the results [30]. The split-half method is a quick and relatively easy way to establish reliability. However, its use is limited to larger questionnaires which only measure one construct. Test item reliability assesses internal consistency through item analysis, where each item on the test is examined to determine how it relates to every other item on the test and to the instrument as a whole [30]. By examining how each item relates to one another, the test item method does not require nearly the length of a test that the split-half method necessitates.

Test-retest reliability (intra-tester and inter-tester) is the degree to which the score remains stable when there is no change in the underlying condition being measured. Test-retest reliability is estimated by measuring individuals two or more times over an interval when the individual's condition is expected to remain stable. The time between the repeat measurements is an important consideration. When the condition being measured is expected to change quickly, the time between repeat measurements should be short. When the condition is not expected to

change rapidly, the time between repeat measurements should be longer [30]. As such, test-retest reliability is not a fixed property of an instrument but rather a degree of measurement consistency when applied to certain populations under particular measurement conditions. Therefore, the population in the study which evaluated test-retest reliability must be the representative of the target population of interest. The reliability coefficients used for test-retest reliability depend on the type of data that is present. For interval or ratio data, the Pearson correlation coefficient and the intraclass correlation coefficient (ICC) are commonly used for normally distributed data. Test-retest reliability for ordinal or nominal data is measured with percent agreement or Cohen's kappa. Reliability values are placed on a common scale of 0–1.0. If a measure is perfect, without error, the reliability is 1.0. If the measures are full of error, the reliability is 0.0. Generally, coefficient values of less than 0.50 indicate poor reliability, 0.50–0.75 represent moderate reliability, and those greater than 0.75 suggest good reliability [30]. In clinical trials, to ensure the valid interpretation of the findings, measures should generally exceed 0.90. However, acceptable reliability is often a judgement call that depends on the knowledge of how precise the measurements must be in order to be used in a meaningful way [30].

In clinical testing, measurements are rarely perfectly reliable [30]. In addition to the limitations of measurement instruments, human subject research adds a level of inconsistency that must be acknowledged [30]. The standard error of measurement (SEM) is a measure of precision that is used to determine such limitations. The SEM estimates how well repeated measures are distributed around a true score. This standard error is directly related to the reliability of a test, with a larger SEM being associated with lower reliability and less precision in the scores obtained.

Precision of the measurement can also be assessed via confidence intervals. A confidence interval gives an estimated range of values, which is likely to obtain the true score for a number of variations in the population [30]. The specific

limits of the confidence interval are determined by the variability in the data as well as the level of confidence the researcher wants to assign to the point estimate [30]. Confidence intervals typically are presented with 95% confidence but can also be reported as 90 or 99%. The confidence interval is reflective of the SEM. The traditionally used 95% confidence interval is determined by multiplying the SEM by 1.96. Likewise, the 90% and 99% confidence intervals are calculated by multiplying the SEM by 1.64 and 2.58, respectively.

The minimal detectable change (MDC) is an absolute measure of reliability or error and is used to determine the threshold for true change in a measure, i.e. what amount of change must be seen to be sure that the difference is not related to measurement error of the instrument [3, 19]. When interpreting scores on an outcome measure, knowing the minimal detectable change is important to ensure the change is not the result of the measurement error alone. The MDC can be determined with knowledge of the SEM and can be applied to findings in clinical research trials.

15.3.2 Validity

Traditionally, validity describes the degree to which an instrument measures what it is intended

to measure. Validity emphasizes the objectives of a test and the ability to make inferences from the associated measurements. *In essence, validity determines what you are able to do with the test results* [30]. Implied in validity is that the measurement has an acceptable level of error or is reliable. Inaccurate or unreliable measurements cannot provide meaningful measurements [30]. Establishing validity of an outcome measure is not as straightforward as establishing reliability [30]. Often there is no obvious manner to determine if an outcome is measuring exactly what it is intending to measure. Therefore, the researcher must determine if the measure has enough validity to be utilized in research and practice through various means.

Validation procedures are based on the type of evidence that can be provided to determine an outcome's validity [30]. Content, criterion-related, and construct validity are essential elements that a researcher must determine to some degree before using an outcome measure.

Content validity is the degree to which items on the instrument adequately reflect the content domain that is being measured. Specifically, content validity addresses the question "are all important item content included on the instrument and all irrelevant item content excluded?" Content validity is useful with questionnaires and inventories [30].

Fact Box 15.1: Reliability

	Description	Statistic used
Internal consistency	The extent to which items measure the same characteristic	Alpha
Test-retest	The extent to which an instrument is able to measure a variable with consistency Not a fixed property of an instrument Degree of measurement consistency when applied to certain populations under particular measurement conditions	Pearson r , ICC Percent agreement, Cohen's kappa
SEM	Interprets internal consistency and test-retest reliability. Measure response stability estimates the standard error in repeated scores	$SEM = SD\sqrt{(1-r_{xx})}$ 95% CI = score \pm (1.96*SEM) 99% CI = score \pm (2.58*SEM)
MDC ₉₅	Minimal amount of change that represents a true change from a change due to variability	$MDC = 1.96 * SEM * \sqrt{2}$

Criterion-related validity is the degree to which the score on the instrument reflects the current or future standing on a gold standard. Essentially, this is an issue that is related to prognosis and how well the score predicts the status of the individual. When appropriate criterion validity is present, the outcomes obtained from one test could be used as a substitute for an established gold standard [30]. There are two standard sub-forms of criterion validity. One sub-form of criterion validity is concurrent validity, which establishes validity when two measures are obtained at a similar time. Establishing concurrent validity is important when one test is considered more efficient, economical, or practical compared to the established gold standard [30]. The other form of criterion validity is predictive validity, which specifically establishes that the outcome on one test can be used to predict the score of the criterion test. Strong criterion validity allows the researcher or clinician to utilize the more efficient and generalizable outcome while acknowledging the similarities to the gold standard.

Finally, construct validity is the degree to which scores on the instrument reflect the under-

lying construct that one is intending to measure. Construct validity requires one to demonstrate hypothesized relationships with other measures of the construct.

15.3.3 Responsiveness

Responsiveness possesses two major elements, internal and external responsiveness [13]. Internal responsiveness is the degree to which the score changes as the underlying condition that is being measured by the scale changes [13]. Essentially it is the instrument’s ability to detect change over time. External responsiveness reflects the extent to which in a measure corresponds to actual changes in a reference measure of health status [13]. With external responsiveness, the measure is not of primary interest but rather the change in the external health status standard [13]. In contrast to internal responsiveness, external responsiveness will depend on the choice of the reference health standard and not on the treatments under investigation [13].

There is a lack of consensus on the appropriate statistic for assessing responsiveness; thus

Fact Box 15.2: Validity

	Description	Determination
Content validity	<ul style="list-style-type: none"> • Degree to which the outcomes adequately reflect the content domain that is being measured • Demands that the outcome is free from the influence of factors that are irrelevant to the purpose of the measurement 	<ul style="list-style-type: none"> • Subjective process with no statistical tests to assess • Determined by expert panel, requiring several revisions
Criterion-related	<ul style="list-style-type: none"> • Reflects the ability of one test to predict the results on another (gold standard) test • The most practical approach to validity testing • Often separated into: <ul style="list-style-type: none"> – Concurrent validity – Predictive validity 	<ul style="list-style-type: none"> • Determined through correlation estimates between the measures
Construct validity	<ul style="list-style-type: none"> • The degree to which scores reflect the underlying constructs that it is measuring • The degree to which a test measures what it claims 	<ul style="list-style-type: none"> • The ability to identify and test theory behind construct is essential • Difficult to ever fully achieve: <ul style="list-style-type: none"> – Estimates of sensitivity – Estimates of specificity – ROC curve analyses – Correlation analyses – Regression models

more than one statistic is often reported [13]. As the patient's condition improves or deteriorates, the score on the measure should change in a similar manner. However, there are a number of factors that can affect the magnitude of change for a given instrument. Factors that affect responsiveness include the patient group (i.e. an acute versus a chronic condition), the type of treatment, the timing of the data collection, and the construct of change [2]. Responsiveness statistics include the effect size, the standardized response means (SRM), the minimally clinically important difference (MCID), and the patient acceptable symptom state (PASS) [2].

The effect size relates change to the standard deviation of the initial scores and is expressed in standard deviation units. It is a way of quantifying the extent of change without confounding it

with sample size. An effect size of 0.5 implies the average change score is equal to one-half of the standard deviation of the initial scores. A general interpretation of effect sizes is that small effect sizes are generally on the magnitude of 0.20, medium are 0.50, and large are around 0.80 [5].

The SRM is an alternative to the effect size and is used to gauge the responsiveness of instruments to actual clinical change. The SRM is determined by dividing the mean change score by the standard deviation of the change score. In clinical research and in determining which outcome measure to use, the measure with the larger SRM will be more able to detect clinical change [1, 23].

The MCID [16] was first described in order to better determine if statistically significant change in an outcome measure also had clinical signifi-

Fact Box 15.3: Responsiveness

	Description	Determination
Effect size	<ul style="list-style-type: none"> Relates change to the standard deviation of the initial scores Expressed in standard deviation units Manner in which to quantify the extent of change without confounding it with sample size Places more emphasis on the size of the effect than statistical significance 	$\text{Effect size} = \left([\bar{x}_1] - [\bar{x}_2] \right) / \text{SD}$
SRM	<ul style="list-style-type: none"> Reflects variability of change scores Provides an estimate of change in the measure, standardized relative to the between patient variability in change scores Removes dependence on sample size Low level of variability in change scores in relation to mean change will have a large SRM value [13] 	$\text{SRM} = \left(\overline{x_2 - x_1} \right) / \text{SD} \left(\overline{x_2 - x_1} \right)$
MCID	<ul style="list-style-type: none"> Smallest amount of change that is identified as clinically important Would dictate a change in patient management 	<ul style="list-style-type: none"> Numerous methods to determine Requires a patient-centred anchor May vary for different populations under different conditions
PASS	<ul style="list-style-type: none"> Highest level of symptom beyond which the patient considers themselves well An absolute level of wellbeing Relatively stable values 	<ul style="list-style-type: none"> Requires anchoring question Use ROC curve analyses to determine cut-point
Ceiling effects	<ul style="list-style-type: none"> All scores cluster at or near the maximum score or best possible outcome Restricts variability as ceiling of test is too low 	<ul style="list-style-type: none"> Frequency of highest possible score achieved by $\geq 15\%$ of subjects
Floor effects	<ul style="list-style-type: none"> All scores cluster at or near the minimum score or worst possible outcome Restricts variability as floor of test is too high 	<ul style="list-style-type: none"> Frequency of lowest possible score achieved by $\geq 15\%$ of subjects

cance or significance to the patient [6]. Given this context, the MCID represents the smallest difference in a score which the patient perceives as beneficial. Unfortunately, there are a number of methods to calculate the MCID, and no standard method has been identified. Therefore, values of MCIDs can have a large amount of variation leading to a number of problems with interpretation [6]. One constant in determining the MCID is the need for a patient-centred anchor, such as the global rating of change in order to determine when patient perceived benefit has taken place. In essence, the MCID functions as a measure of responsiveness of a given instrument. However, at times the MCID may be less reflective of the responsiveness of the instrument and more reflective of the treatment itself [6].

The PASS is defined as the highest level of the symptom beyond which the patient considers themselves well [20]. Like the MCID, the PASS also requires an anchoring question to identify the cut-off. The anchoring question “Taking into account all of the activities you have during your daily life, do you consider that your current state satisfactory?” has simple response options of yes or no. PASS cut-points appear to be relatively stable over time and are not strongly influenced by age or gender [20]. PASS can be determined for both patient-reported outcomes and clinical measures such as pain and movement ability.

When the average scores for the group of interest represent different points on a measurement instrument and the instrument is not able to register equal intervals across the full range of measurements, ceiling and floor effects occur [30]. Ceiling effects technically occur when the outcome is equal across the range of treatment groups, and all groups cluster at or near the best possible outcome level. Conversely, floor effects occur when performance is nearly as bad as possible in all groups or at the low of the instrument scale. Outcome instruments with either ceiling or floor effects render the instrument unable to discriminate between patients at either extreme of the scale [32]. Generally, ceiling or floor effects are said to be present in an instrument when 15% or more of the respondents achieve the best or worst level of the score [24, 28].

Clinical Vignette

When designing a research study to examine the outcomes of three separate surgical procedures for reconstruction of the anterior cruciate ligament with comparable post-operative rehabilitations, the investigators needed to identify the most important outcome measures. Given that all three surgical techniques were established and the purpose of the study had to deal with the overall patient outcome and improvement in symptoms, function, and general quality of life, the outcome measures selected were required to reflect those domains. Performance-based measures such as range of motion, strength, and proprioception were selected to examine the effect on impairments and symptoms and were based on factors that are considered important fundamental aspects that are related to outcomes. Patient-reported outcomes included self-reported pain, condition-specific and region-specific questionnaires, and general health status indices to examine overall quality of life. When selecting the outcome measures, the investigators researched measures that were found to have sufficient reliability and high validity to their patient population and that were responsive to change. Therefore, the International Knee Documentation Committee Subjective Knee Form (IKDC-SKF) [14] was selected over the WOMAC [4] as it has superior validity for the population undergoing ACL reconstruction. At the termination of the study, the investigators were able to draw conclusions on multiple aspects of the effect of the intervention on impairments, activity and participation limitations, and general quality of life. Since the outcomes were methodologically sound, the results had good generalizability to the population they were studying and allowed comparison with similar research trials available in the literature.

Applying the concepts of reliability, validity, and responsiveness to evaluate a measure of health-related quality of life is essential when selecting outcomes for clinical trials. Whether an outcome measure is performance-based or patient-reported, general or specific, all measures must be assessed for acceptable levels of reliability in the measurement, validity in the construct that it measures, and its ability to detect change that may occur with intervention. More generally, outcome measures should match the level of intervention that is being provided in the clinical trial, with clinical trials addressing impairments, activity limitations, or participation restrictions having outcome measures that correspondingly assess issues at the impairment, functional, and societal levels.

Take-Home Message

- To comprehensively assess results of clinical trials, clinical outcome measures should include those that promote comparison to the population as a whole, as well as measures that are distinctive to the condition and population of interest via region- or disease-specific assessments.
- Both performance-based and patient-reported outcomes are useful, with patient-reported outcomes being a key indicator of patient-centered responses.
- Consideration of the psychometric properties of reliability, validity, and responsiveness of the measure as they relate to the condition of interest is essential when selecting clinical outcomes to ensure that there is methodological acceptability in the measures.

15.4 Resources/Websites

<http://www.orthopaedicscore.com>
www.rehabmeasures.org
https://www.aaos.org/Quality/Performance_Measures/Patient_Reported_Outcome_Measures/?ssopc=1
<http://www.ptnow.org/tests-measures>

References

1. Evidence-based rehabilitation: a guide to practice. 2nd ed. Thorofare: SLACK Incorporated; 2008.
2. Beaton DE. Understanding the relevance of measured change through studies of responsiveness. *Spine*. 2000;25(24):3192–9.
3. Beaton DE, Bombardier C, Katz JN, Wright JG, Wells G, Boers M, et al. Looking for important change/differences in studies of responsiveness. OMERACT MCID Working Group. Outcome measures in rheumatology. Minimal clinically important difference. *J Rheumatol*. 2001;28(2):400–5.
4. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol*. 1988;15(12):1833–40.
5. Cohen J. Statistical power analysis. *Curr Dir Psychol Sci*. 1992;1(3):98–101.
6. Cook CE. Clinimetrics corner: the minimal clinically important change score (MCID): a necessary pretense. *J Man Manip Ther*. 2008;16(4):E82–E3.
7. Fairbank JC, Couper J, Davies JB, O'Brien JP. The Oswestry low back pain disability questionnaire. *Physiotherapy*. 1980;66(8):271–3.
8. Guralnik JM, Branch LG, Cummings SR, Curb JD. Physical performance measures in aging research. *J Gerontol*. 1989;44(5):M141–M6.
9. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Qual Life Res*. 2009;18(7):873–80.
10. Hefford C, Abbott JH, Arnold R, Baxter GD. The patient-specific functional scale: validity, reliability, and responsiveness in patients with upper extremity musculoskeletal problems. *J Orthop Sports Phys Ther*. 2012;42(2):56–65.
11. Horn KK, Jennings S, Richardson G, Van Vliet D, Hefford C, Abbott JH. The patient-specific functional scale: psychometrics, clinimetrics, and application as a clinical outcome measure. *J Orthop Sports Phys Ther*. 2012;42(1):30–42.
12. Hudak PL, Amadio PC, Bombardier C, Beaton D, Cole D, Davis A, et al. Development of an upper extremity outcome measure: the DASH (Disabilities of the Arm, Shoulder, and Hand). *Am J Ind Med*. 1996;29(6):602–8.
13. Husted JA, Cook RJ, Farewell VT, Gladman DD. Methods for assessing responsiveness: a critical review and recommendations. *J Clin Epidemiol*. 2000;53(5):459–68.
14. Irrgang JJ, Anderson AF, Boland AL, Harner CD, Kurosaka M, Neyret P, et al. Development and validation of the international knee documentation committee subjective knee form. *Am J Sports Med*. 2001;29(5):600–13.

15. Irrgang JJ, Snyder-Mackler L, Wainner RS, Fu FH, Harner CD. Development of a patient-reported measure of function of the knee. *J Bone Joint Surg Am*. 1998;80(8):1132–45.
16. Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials*. 1989;10(4):407–15.
17. Kirkley A, Alvarez C, Griffin S. The development and evaluation of a disease-specific quality-of-life questionnaire for disorders of the rotator cuff: the Western Ontario Rotator Cuff Index. *Clin J Sport Med*. 2003;13(2):84–92.
18. Kirshner B, Guyatt G. A methodological framework for assessing health indices. *J Clin Epidemiol*. 1985;38(1):27–36.
19. Kovacs FM, Abaira V, Royuela A, Corcoll J, Alegre L, Tomás M, et al. Minimum detectable and minimal clinically important changes for pain in patients with nonspecific neck pain. *BMC Musculoskelet Disord*. 2008;9(1):43.
20. Kvien TK, Heiberg T, Hagen KB. Minimal clinically important improvement/difference (MCII/MCID) and patient acceptable symptom state (PASS): what do these concepts mean? *Ann Rheum Dis*. 2007;66(Suppl 3):iii40–1.
21. Latham NK, Mehta V, Nguyen AM, Jette AM, Olarsch S, Papanicolaou D, et al. Performance-based or self-report measures of physical function: which should be used in clinical trials of hip fracture patients? *Arch Phys Med Rehabil*. 2008;89(11):2146–55.
22. Leggin B, Iannotti J. *Shoulder outcome measurement. Disorders of the shoulder: diagnosis and management*. Philadelphia: Lippincott Williams & Wilkins; 1999. p. 1024–40.
23. Liang MH, Lew RA, Stucki G, Fortin PR, Daltroy L. Measuring clinically important changes with patient-oriented questionnaires. *Med Care*. 2002;40(4):II-45–51.
24. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res*. 1995;4(4):293–307.
25. McHorney CA, Ware JE Jr, Raczek AE. The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care*. 1993;31:247–63.
26. Nagi SZ. A study in the evaluation of disability and rehabilitation potential: concepts, methods, and procedures. *Am J Public Health Nations Health*. 1964;54(9):1568–79.
27. Noyes FR, Barber SD, Mooar LA. A rationale for assessing sports activity levels and limitations in knee disorders. *Clin Orthop*. 1989;(246):238–49.
28. Paulsen A, Odgaard A, Overgaard S. Translation, cross-cultural adaptation and validation of the Danish version of the Oxford Hip Score: assessed against generic and disease-specific questionnaires. *Bone Joint Res*. 2012;1(9):225–33.
29. Porter ME. What is value in health care? *N Engl J Med*. 2010;363(26):2477–81.
30. Portney L, Watkins M. *Foundation of clinical research: application to practice*. Norwalk: Appleton & Lange; 1993.
31. Rankin G, Stokes M. Reliability of assessment tools in rehabilitation: an illustration of appropriate statistical analyses. *Clin Rehabil*. 1998;12(3):187–99.
32. Stucki G, Liang M, Stucki S, Katz J, Lew R. Application of statistical graphics to facilitate selection of health status measures for clinical practice and evaluative research. *Clin Rheumatol*. 1999;18(2):101–5.
33. Tegner Y, Lysholm J. Rating systems in the evaluation of knee ligament injuries. *Clin Orthop*. 1985;(198):42–9.
34. Vernon H, Mior S. The neck disability index: a study of reliability and validity. *J Manipulative Physiol Ther*. 1991;14(7):409–15.
35. Ware JE Jr, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996;34(3):220–33.
36. *International Classification of Functioning, Disability and Health (ICF)* [press release]. Geneva: World Health Organization; 2001.

Advances in Measuring Patient-Reported Outcomes: Use of Item Response Theory and Computer Adaptive Tests

Andrew D. Lynch, Adam J. Popchak,
and James J. Irrgang

16.1 Introduction to Item Response Theory

Most clinical outcomes do not fall into discrete categories (e.g., alive or dead; torn or intact), but rather are measured on a continuum (e.g., physical function, pain). This is especially true for patient-reported outcomes (PROs), which classically place an individual on the continuum of the underlying construct being measured via scores on fixed-length surveys. An instrument calibrated with item response theory uses individual items to estimate the location on the continuum or latent trait being measured by the set of items [4]. Using multiple items from a single bank of items which all measure the same construct (e.g., mobility or upper extremity function) improves the precision of that measurement.

Item response theory (IRT) is a modern measurement method that can help to eliminate redundant items and ensure that an item bank is unidimensional [8]. The underlying premise of IRT is that the performance of a person on an item can be modeled by the characteristics of the person and the item. Item response theory is a family of mathematical models that explains how individuals at different ability levels should

respond to an item. As the ability of an individual increases, the probability of choosing a “correct” response to the item increases.

For example, individuals with higher functional ability, such as athletes, will have a higher probability than nonathletes of responding positively to an item “Can you run one mile?”. Therefore, this item will be more likely administered to someone with good functional ability. An in-depth description of the mathematics of the models underlying IRT is beyond the scope of this chapter. The interested reader is directed to Hays [13].

IRT calibration assumes that the item bank is unidimensional—that it measures a single latent trait and can therefore be expressed as a single score [13, 17]. In contrast, consider a measure that asks questions about both physical pain and depression—an individual can have significant physical pain but not be depressed or can be depressed without having physical pain. A single score on this hypothetical measure is difficult to interpret because it measures more than one dimension of health-related quality of life.

Many traits are measured on extremely broad spectrum. Mobility related to physical function is an excellent example of a broad latent trait. At the low end of the mobility spectrum, an individual may be bedridden, requiring assistance to roll over. At the high end, a world-class decathlete has very high levels of mobility. Between these two individuals lies a broad continuum which

A. D. Lynch (✉) · A. J. Popchak · J. J. Irrgang
Departments of Physical Therapy and Orthopaedic
Surgery, University of Pittsburgh,
Pittsburgh, PA, USA
e-mail: adl45@pitt.edu

must be considered. Any number of questions exist, which can place an individual somewhere along the continuum from immobility to very high levels of mobility. To perfectly place the individual on the continuum, a PRO should contain a range of questions that measure the full spectrum of the trait including the lowest and highest end of the trait. In our mobility example, this would include asking questions about rolling over in bed and being able to run the 400-m hurdles—and everything in between. A measure with a large number of items would be incredibly long and burdensome to complete and just as burdensome to interpret. However, a scale with only ten items may be subject to floor and ceiling effects as noted in Chap. 15.

IRT can be used to calibrate a set of items (i.e., an item bank) that will measure the extremes of function. Once the items are calibrated, this can be used to measure an individual on the continuum without having to administer a large number of items [19]. Each response to an item gives an estimate of an individual's *location* on the continuum of the latent trait.

An item can be mathematically described by the item characteristic curve (ICC). The ICC represents the “difficulty” and slope of an item. The difficulty relates to the level of latent trait being measured by the item. For example, an item difficulty of 0.5 provides the most information about someone whose level of the latent trait is 0.5 logits above average. A greater slope indicates that an item is more discriminating around the level of difficulty for an item (i.e., a small change in the latent trait is likely to change the response) [13]. To select the best item for an individual, we should try to match the item difficulty as closely as possible to the individual's level of the latent trait. Then we should select the item that has the greatest slope (i.e., the item with the greatest discrimination).

When items are calibrated using item response theory (IRT), a scaling factor and difficulty rating are assigned to each item, so that the response on any single item can be compared to other items in the bank [20]. Measures calibrated with IRT can locate an individual on the functional scale using a subset of items in a calibrated item bank. Therefore, direct comparisons of individuals are possible using two completely different item sets.

16.2 Computer Adaptive Test (CAT) Technology

Computer adaptive testing (CAT) algorithms choose items that will provide the most information about an individual (i.e., items that will give the best estimate of an individual's position on the latent trait) [4, 13]. A great benefit of an item bank that is calibrated using IRT is the ability to accurately predict the response to many items based on the individual response to a few items. It is intuitive that an individual who responds that he cannot walk 1 mile will also not be able to run 5 miles. If a clinician is interviewing her patient, she would not ask follow-up questions that do not make sense. The clinician would ask questions that provide the most value and information about the function of the individual. Administering an item bank using CAT methods uses algorithms to accomplish the same end goal. The CAT algorithms improve the efficiency of item administration by presenting items that are relevant to the individual. Generally, the position on the scale can be determined after administration of four to eight items. This greatly reduces respondent burden compared to traditional general measures [e.g., Medical Outcomes Study Short Form 36 (SF-36—36 items)] or region-specific items [e.g., Knee Injury and Osteoarthritis Outcome Score (KOOS)—42 items].

To illustrate the concept of a CAT, consider the following example. Based on the age and sex of an individual, the initial question administered to a 25-year-old female might be “how much difficulty do you have walking a mile?”. If the individual indicates to have no difficulty walking a mile, the computer algorithm would bypass “easier” items, such as “can you walk a block” and would administer a more “difficult” item, “how much difficulty do you have running a mile?”. After each item is administered, the individual's level of physical function is re-estimated, and an item that would provide the most information given the individual's current estimate of function is administered. The process continues iteratively until a predetermined number of items are administered or the level of function is estimated with a pre-specified level of precision.

Item banks are designed to measure the most common presentations of the latent trait. As an

example, the Patient-Reported Outcome Measurement Information System (PROMIS) Physical Function Item Bank is designed to measure the physical function of individuals in the general population and to distinguish between someone who is functioning “normally” and someone who is limited in some fashion. It was not designed to accurately measure the function of either elite athletes or those who are nearly bedridden. One way to assess the measurement capability of a CAT is to graphically look at a representation of the location vs. precision. We graph the location of the score along the x -axis and the precision associated with that score on the y -axis. Generally, a U-shaped graph is seen with a range of scores that are associated with a precision (standard error) of less than 3.3 in the center of the graph and less precise scores at either end.

However, some measures do not follow this U-shape, as can be seen in the ability of the PROMIS Pain Interference CAT and Physical Function CAT in Fig. 16.1. Each measure is associated with good precision (i.e., standard error of less than 3.3 for scores between -1 to 1). However, for Pain Interference t -scores less than 40 (scores indicating that pain does interfere with daily function), there is poor precision ($SE > 5$). Similarly, for Physical Function scores greater than 55 , precision is

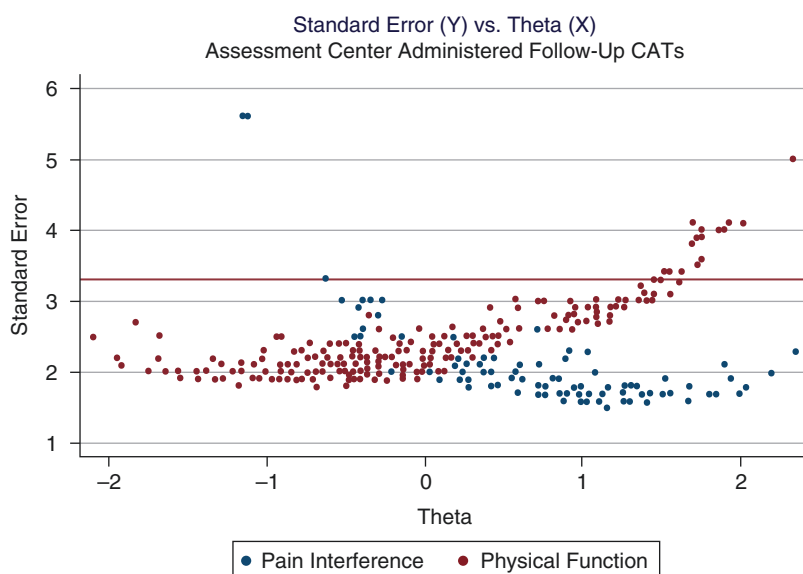
again poor. Considering these results, the Pain Interference CAT may not be suitable for individuals for whom pain is only a minor inconvenience, and the Physical Function CAT may not be suitable for individuals who function at the highest end of the spectrum. Each of these scales may benefit from replenishment of the item bank to expand the range of the latent trait being measured (see below).

16.3 CAT Administration Parameters

When setting up the CAT parameters, the user can create stopping rules for administration of items. The most intuitive of these stopping rules is the number of items to administer. Generally, the user can state the minimum and maximum number of items to administer, which generally is set at a minimum of 4 items and a maximum of 12 items for the PROMIS instruments.

The second stopping rule involves the overall precision of the location estimate. When estimating the position of the individual on the scale of the latent trait, we want to know how precise that estimate is. Having an estimate that your patient is functioning at the 50th percentile of mobility is helpful if the 90% confidence interval (i.e., the

Fig. 16.1 Sample data (previously unpublished) for PROMIS CAT scores in individuals with orthopedic knee conditions



true location) is between the 47th and 53rd percentiles. However, if the 90% confidence interval ranges from the 30th to the 70th percentiles, we are much less confident in our estimate. Therefore, the administration setup often allows the user to identify a level of precision with which they are comfortable for the location estimate. Depending on the desired use of the data, this can range from two to ten standard error points. There is a trade-off in the number of items for being more precise—with greater required precision, generally more items must be administered.

It is important to understand that the number of items to administer generally overrides the precision requirement. If the precision requirement is

met after three items, but the minimum was set to four, the algorithm will administer a fourth item. If the maximum number of items is administered, but the precision requirement has not been met, the algorithm will stop administering items.

16.4 Interpreting Score Reports from Computer Adaptive Tests

Scores are typically expressed as a *t*-score. These standardized scores are used when a large, normative data set is available and encourages comparison of the individual to the population average as is available for the PROMIS measures [6, 16]. When expressed as a *t*-score, the expected mean is 50, and the expected standard deviation is 10. Therefore, we can expect that 68% of individuals will score between 40 and 60, 90% between 30 and 70, and 99.7% between 20 and 80. We also then know that a difference of 10 points on a *t*-score between individuals represents about one standard deviation.

As mentioned in Chap. 15, psychometric properties of an instrument are not fixed. Reliability, responsiveness, and patient acceptable symptom state may change depending on the population. Prior to choosing a CAT measure, the investigator should assess whether it has been used in a similar population. The PROMIS Physical Function mea-

Fact Box 16.1: Computer Adaptive Testing

- Computer adaptive testing (CAT) uses an algorithm to choose an item from an IRT-calibrated item bank that will provide the most information about an individual. Therefore, two individuals can complete the same CAT-administered outcome measure and respond to different items.
- The score on a CAT for an IRT-calibrated measure considers only the responses to individual items that are administered to the individual and not responses to a complete item bank.
- Therefore, a CAT does not need to administer all items to arrive at a score.
- Because the administration is an estimate of a score, there is an associated standard error with the score. The lower the standard error, the more precise the estimate.
- Most often, a minimum of 4 and maximum of 12 items are administered; once four items have been administered and an acceptable level of precision for the score estimate has been achieved, testing stops. This makes test administration more efficient than administration of a full item bank or fixed-length outcome measure.

Fact Box 16.2: *t*-Score Interpretation

- A *t*-score normalizes an individual score relative to a large, representative population who have also completed the outcome measure.
- On a *t*-score, 50 represents the population average and 10 points is 1 standard deviation.
- Because *t*-scores are normally distributed, 68% of individuals will score between 40 and 60, 90% between 30 and 70, and 99.7% between 20 and 80.
- Both CAT administration and short form administration will provide a *t*-score.

asures have been used routinely in orthopedic populations [1, 2, 12, 15, 18, 21].

16.5 Assessment with Short Forms

Computer adaptive tests obviously require computers and depending on the mode of administration (e.g., Research Electronic Data Capture System (REDCap) or measure-specific web site) an Internet connection. However, there are workarounds for scenarios in which an Internet-connected device capable of administering the measure is not available.

Using the item characteristic curves, it is possible to create static short forms that resemble classic fixed-length patient-reported outcome surveys. A short form has a fixed number of items, to which the patient responds. There are scores associated with each of the items; however, scoring a short form is not as simple as totaling the points on the form or expressing the achieved points as a percentage of the total points possible as is typically done with legacy outcome measures.

After totaling the points from all items, the person scoring the form must refer to a conversion table to arrive at a *t*-score and standard error. Conversion tables are typically available in administration manuals.

Reviewing a conversion table prior to choosing a short form can help to identify if a particular short form will meet the needs of the clinician-researcher. Because each short form “total score” is associated with a *t*-score and a standard error, the clinician-researcher can review the range of *t*-scores that are associated with a particular short form and the associated standard errors. As an example, multiple PROMIS Physical Function Short Forms are available from HealthMeasures.org (e.g., Physical Function 4a, Physical Function 6b, and Physical Function 8b). The maximum *t*-score for each of these measures is 57, 59, and 60.1, respectively; however, each is associated with a standard error of 5.9 or greater. These short forms are all capable of measuring physical function that is above the population average of

50; however, the location estimates for someone who achieves the highest possible score on the form is not a very precise measure. On the other hand, raw summed scores that are not at the maximum are typically associated with a standard error of 3 or less, which is usually regarded as acceptable. Therefore, these forms would have an obvious ceiling effect when measuring physical function at the highest level of function.

In the instance where a broad range of a latent trait must be measured, it is possible to select two versions of the short form—one for the low end of the spectrum and one for the high end. It is up to the clinician to judge which version would best serve the individual patient. In this case, there may be a total of 24 items that could be administered, but any given individual will only be asked to complete half of them to arrive at a *t*-score with an acceptable level of precision. Alternatively, to determine if the high or low version of the short form should be administered, a single screening item that is highly discriminative may be administered first. The response to the item would determine if the high or low version of the short form is administered.

16.6 Expanding and Improving Content Coverage

IRT-calibrated item banks are superior to classically created, fixed-length assessments for measuring at the extremes of function, but the item bank must cover the complete range of the intended trait for precise measurement [9, 11, 14]. The ability to improve the psychometric properties of a measure by adding and calibrating new items—referred to as item bank replenishment—is a significant advantage of IRT-based measures [10]. In the scenario where an item bank is deemed insufficient to measure a certain aspect of function, a *replenishment study* may be completed to augment the item bank. Replenishment studies typically seek to expand the upper or lower boundaries of the item bank to improve the range of measurement; however, it is also conceivable that an item bank may need better coverage in the middle of the spectrum of the latent trait.

In a replenishment study, additional items are identified via comparison to other existing item banks and legacy outcome measures and through interviews with relevant stakeholder groups [3, 4, 7, 10]. For instance, an item bank may be excellent at measuring general mobility, but may not have items that appropriately measure mobility that is assisted by a cane, walker, or wheelchair. In this case, existing measures could be consulted, for example, items, and individuals who routinely get around with assistance may be asked to provide input on what types of items should be included or how the items should be worded. This provides a group of candidate items that may be used to augment the existing bank.

The item bank calibration method is a more direct method to calibrate new items. In this instance, an individual is asked to respond to a large portion of the existing item bank (possibly the entire item bank) and to each new candidate item. Because there is more information with which to calibrate the new items, fewer overall participants are needed. However, each participant assumes a greater burden by responding to a large number of items. Regardless, an item bank is not a completely static entity. It may be augmented and refined through multiple administrations.

16.7 Using CATs in Conjunction with Legacy Measures

Presently, the majority of IRT-calibrated measures that can be administered as CATs are general measures. As an example, the PROMIS has tools to measure fatigue, pain intensity, pain interference, physical function, sleep disturbance, anxiety, depression, and ability to participate in social roles and activities as their primary measures. These are not disease- or condition-specific measures, and therefore, they will not ask specific questions about how a recent rotator cuff tear affects sleep or physical function. If disease-specific or body region-specific information is needed, an additional legacy measure

should be considered in addition to the general measure.

It can be argued that a region-specific measure related to the hip or a condition-specific measure about osteoarthritis gives more information about a particular patient's situation. However, it is not possible to compare scores on a shoulder-specific measure and a knee-specific measure. It is possible to determine how each of those individuals is impacted in general physical function when measured by a general IRT-based CAT.

16.8 Limitations Associated with Computer Adaptive Testing

Legacy patient-reported outcome measures, with a fixed number of items that are always administered, are easy to program into the electronic medical record (EMR) for administration in the clinic. However, CAT-based measures require specific, proprietary algorithms to be programmed into the EMR. This is not yet a standard practice for the EMR; therefore, adoption in clinical practice is difficult and has not yet become routine. Alternatively, research based programs such as the Research Electronic Data Capture (REDCap) System have the PROMIS CATs built into their system. Additionally, the vendors of each program have websites and applications that can administer the CAT-based PROs, usually at a cost.

The CAT-based PRO may also lack specificity to the patient or clinician [5]. Because it is a general measure of a specific trait, it contains general items. This is unlike legacy PROs which may be joint or region specific and therefore contain items about a specific injury, symptom, or function. Some individuals comment that the CAT-based PROs are not specific to their current situation, and therefore they do not see value in completing it. To address and overcome this concern, it is imperative that the clinician look at the PRO and discuss the results with the individual.

Clinical Vignette

In an orthopedic clinic, a group of partners provide care for a variety of orthopedic presentations. They are curious to understand how each of their patients are functioning on the continuum of physical function. Therefore, they decide to administer the PROMIS Physical Function CAT to each patient at each visit to track function and progress. An example of the *t*-scores and standard errors are described below for Patient A, who presented with an ACL tear requiring reconstruction and for Patient B, who presented with a degenerative meniscus tear requiring debridement.

Because the items in the PROMIS Physical Function Item Bank are all calibrated on a single scale, the scores can be directly compared. It does not matter that the individuals did not respond to the same items. We can clearly see that Patient A is doing better in his overall mobility compared to Patient B at all time points. Importantly, we were able to arrive at a score in only four items for most administrations. The only instance which required more than four items was when Patient A

was functioning at the highest level of mobility, where we know measurement precision is poor.

While Patient B may not have achieved the same level of overall mobility, when we consider the Pain Interference scores for this individual, we see an initial trend of increased pain after surgery, but ultimately a reduction of five *t*-score points or one-half of a standard deviation. This is an excellent example of being able to measure two constructs (Physical Function and Pain Interference) in an individual in only eight items. Pain Interference scores were low for Patient A and did not vary over time.

However, because the Physical Function Item Bank does not ask the same questions every time the CAT is administered, we do not directly know what physical limitations are present in either Patient A or Patient B. This would require administration of a knee-joint-specific outcome measure. We may also estimate how an individual would respond to item that were not administered based on responses to other IRT-calibrated items.

	Patient A: ACL reconstruction			Patient B: arthroscopic debridement of degenerative meniscus					
	Physical function <i>t</i> -score	SE	Item count	Physical function <i>t</i> -score	SE	Item count	Pain interference <i>t</i> -score	SE	Item count
Pre-op	47.7	1.9	4	39.1	1.9	4	57.7	1.7	4
3 m	44.7	2.1	4	38.3	1.8	4	65.7	1.7	4
6 m	47.2	2	4	40.6	1.9	4	57.3	1.7	4
12 m	54.6	2.4	4	45.7	1.9	4	52.8	1.8	4
24 m	70.3	4.1	12	48.7	1.9	4	52.6	1.9	4

Take-Home Message

- The use of IRT and CAT to administer patient-reported outcomes allows for more rapid assessment of various aspects of health-related quality of life.
- Widespread use of IRT-based measures will promote comparisons between studies and better meta-analysis; however, CAT administration has not become part of routine practice yet.
- It remains to be seen how IRT- and CAT-administered PROs function in specific patient populations.

16.9 Resources and Web Sites

- PROMIS Resources via Healthmeasures.net—<http://www.healthmeasures.net>
- Activity Measure for Post-Acute Care—<http://am-pac.com/category/home/>

References

1. Beleckas CM, Padovano A, Guattery J, Chamberlain AM, Keener JD, Calfee RP. Performance of patient-reported outcomes measurement information system (PROMIS) upper extremity (UE) versus physical function (PF) computer adaptive tests (CATs) in upper extremity clinics. *J Hand Surg Am.* 2017;42:867–74. <https://doi.org/10.1016/j.jhsa.2017.06.012>.
2. Bernholt D, Wright RW, Matava MJ, Brophy RH, Bogunovic L, Smith MV. Patient reported outcomes measurement information system scores are responsive to early changes in patient outcomes following arthroscopic partial meniscectomy. *Arthroscopy.* 2018;34:1113–7. <https://doi.org/10.1016/j.arthro.2017.10.047>.
3. Bruce B, Fries J, Lingala B, Hussain YN, Krishnan E. Development and assessment of floor and ceiling items for the PROMIS physical function item bank. *Arthritis Res Ther.* 2013;15:R144. <https://doi.org/10.1186/ar4327>.
4. Bruce B, Fries JF, Ambrosini D, Lingala B, Gandek B, Rose M, Ware JE Jr. Better assessment of physical function: item improvement is neglected but essential. *Arthritis Res Ther.* 2009;11:R191. <https://doi.org/10.1186/ar2890>.
5. Cella D, Chang CH. A discussion of item response theory and its applications in health status assessment. *Med Care.* 2000;38:II66–72.
6. Cella D, et al. The patient-reported outcomes measurement information system (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005–2008. *J Clin Epidemiol.* 2010;63:1179–94. <https://doi.org/10.1016/j.jclinepi.2010.04.011>.
7. DeWalt DA, Rothrock N, Yount S, Stone AA, PROMIS Cooperative Group. Evaluation of item candidates: the PROMIS qualitative item review. *Med Care.* 2007;45:S12–21. <https://doi.org/10.1097/01.mlr.0000254567.79743.e2>.
8. Fries JF, Bruce B, Cella D. The promise of PROMIS: using item response theory to improve assessment of patient-reported outcomes. *Clin Exp Rheumatol.* 2005;23:S53–7.
9. Haley SM, Ni P, Hambleton RK, Slavin MD, Jette AM. Computer adaptive testing improved accuracy and precision of scores over random item selection in a physical functioning item bank. *J Clin Epidemiol.* 2006;59:1174–82. <https://doi.org/10.1016/j.jclinepi.2006.02.010>.
10. Haley SM, Ni P, Jette AM, Tao W, Moed R, Meyers D, Ludlow LH. Replenishing a computerized adaptive test of patient-reported daily activity functioning. *Qual Life Res.* 2009;18:461–71. <https://doi.org/10.1007/s11136-009-9463-5>.
11. Haley SM, Ni P, Ludlow LH, Fragala-Pinkham MA. Measurement precision and efficiency of multidimensional computer adaptive testing of physical functioning using the pediatric evaluation of disability inventory. *Arch Phys Med Rehabil.* 2006;87:1223–9. <https://doi.org/10.1016/j.apmr.2006.05.018>.
12. Haskell A, Kim T. Implementation of patient-reported outcomes measurement information system data collection in a private orthopaedic surgery practice. *Foot Ankle Int.* 2018;39:517–21. <https://doi.org/10.1177/1071100717753967>.
13. Hays RD, Morales LS, Reise SP. Item response theory and health outcomes measurement in the 21st century. *Med Care.* 2000;38:II28–42.
14. Jette AM, Haley SM. Contemporary measurement techniques for rehabilitation outcomes assessment. *J Rehabil Med.* 2005;37:339–45. <https://doi.org/10.1080/16501970500302793>.
15. Lee AC, Driban JB, Price LL, Harvey WF, Rodday AM, Wang C. Responsiveness and minimally important differences for 4 patient-reported outcomes measurement information system short forms: physical function, pain interference, depression, and anxiety in knee osteoarthritis. *J Pain.* 2017;18:1096–110. <https://doi.org/10.1016/j.jpain.2017.05.001>.
16. Liu H, Cella D, Gershon R, Shen J, Morales LS, Riley W, Hays RD. Representativeness of the patient-reported outcomes measurement information system internet panel. *J Clin Epidemiol.* 2010;63:1169–78. <https://doi.org/10.1016/j.jclinepi.2009.11.021>.
17. McHorney CA, Cohen AS. Equating health status measures with item response theory: illustrations with functional status items. *Med Care.* 2000;38:II43–59.
18. Minoughan CE, Schumaier AP, Fritch JL, Grawe BM. Correlation of PROMIS physical function upper extremity computer adaptive test with American shoulder and elbow surgeons shoulder assessment form and simple shoulder test in patients with shoulder arthritis. *J Shoulder Elb Surg.* 2018;27:585–91. <https://doi.org/10.1016/j.jse.2017.10.036>.
19. Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE Jr. The PROMIS physical function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. *J Clin Epidemiol.* 2014;67:516–26. <https://doi.org/10.1016/j.jclinepi.2013.10.024>.
20. Tao W, Haley SM, Coster WJ, Ni P, Jette AM. An exploratory analysis of functional staging using an item response theory approach. *Arch Phys Med Rehabil.* 2008;89:1046–53. <https://doi.org/10.1016/j.apmr.2007.11.036>.
21. Zdziarski-Horodyski L, et al. An integrated-delivery-of-care approach to improve patient reported physical function and mental wellbeing after orthopaedic trauma: study protocol for a randomized controlled trial. *Trials.* 2018;19:32. <https://doi.org/10.1186/s13063-017-2430-5>.

Part III

Basics in Statistics: Statistics Made Simple!

Stephan Bodkin, Joe Hart, and Brian C. Werner

17.1 Introduction

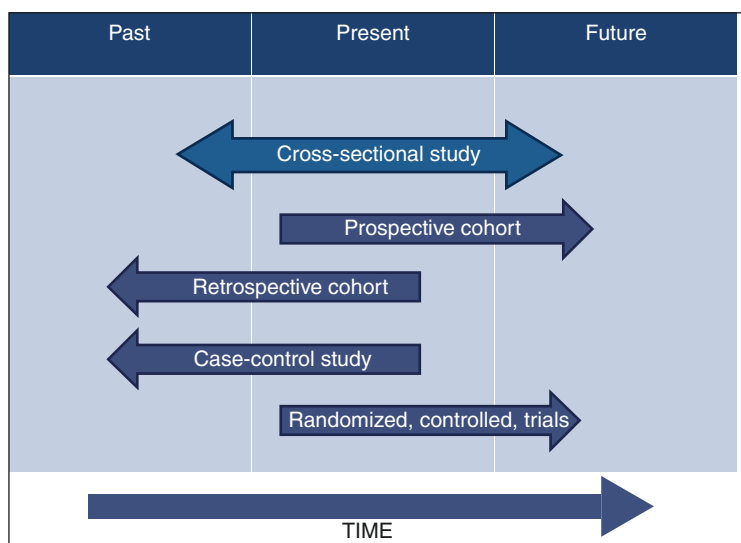
17.1.1 Common Research Designs

Research studies can be either *prospective* or *retrospective* in nature. Retrospective studies are performed by identifying a target population and making observations backwards in time. Common retrospective studies are carried out

through patient questionnaires and medical chart reviews. Prospective studies are those that make observations forward in time. A longitudinal study, which is prospective in nature, makes repeated observations of the same variables over time (Fig. 17.1).

Observational studies, often referred to as *descriptive studies*, are those in which no treatment or intervention is administered to the sub-

Fig. 17.1 Timing of common research designs. Cross-sectional studies involve collection of similar subjects at different time points. Prospective studies identify cases of interest and follow them forward in time. Retrospective studies will identify cases and observe characteristics backwards in time; common examples of retrospective studies are chart reviews



S. Bodkin · J. Hart · B. C. Werner (✉)
 Department of Orthopaedic Surgery, University of
 Virginia, Charlottesville, VA, USA
 e-mail: BCW4X@hscmail.mcc.virginia.edu

jects. Due to this, no casual effect can be established. Observational studies can be either retrospective or prospective in design. Observational studies are often inexpensive and take less time to complete compared to intervention studies; however, as discussed later in this chapter, their results possess a lower level of evidence. These studies are commonly performed to develop hypotheses on which larger scale studies can be performed.

17.1.2 Observational Research Designs

Case reports represent the reporting of outcomes or descriptions of one patient. The patient usually presents a unique injury or illness which may be difficult to collect and study in larger numbers. There can be no attempt in data analysis with case report studies. A *case series* represents the reporting of clinical outcomes of multiple patients with the same injury, illness, or treatment management. The number of subjects in a case series can range from a few to a couple dozen depending on the prevalence of the characteristic of interest. Both case reports and case series do not have a control group and do not utilize an experimental design. They are not designed to provide information about frequency or distribution of the pathology of interest.

Cohort studies, which can be either prospective or retrospective, are aimed to identify and study a group of subjects with similar characteristics, whether that is a pathology or treatment. Common data collected within a cohort study includes subjective (patient-reported) and objective (clinician-measured) outcomes after injury, surgery, or the characteristic of interest. Statistical tests can be used to find associations and differences among groups and can be interpreted for clinical importance. The difference between a prospective cohort study and randomized control trial is the lack of randomization of a treatment or intervention.

Case-control studies are a form of cohort study; however, comparisons are made between groups of subjects based on an outcome rather than an exposure. Data received from the study

cohort of interest are compared to those of a control group.

Cross-sectional studies involve data observations of a particular study cohort at one time point with no follow-up. These studies are thought of being a “snapshot” of the cohort of interest. Though not as strong as longitudinal studies, with collection of similar subjects at different time points post-diagnosis, postsurgery, etc., the researcher may get an idea of how the characteristic changes over time. A cross-sectional study is useful to describe the frequency of a characteristic of interest to make associations with other variables over time.

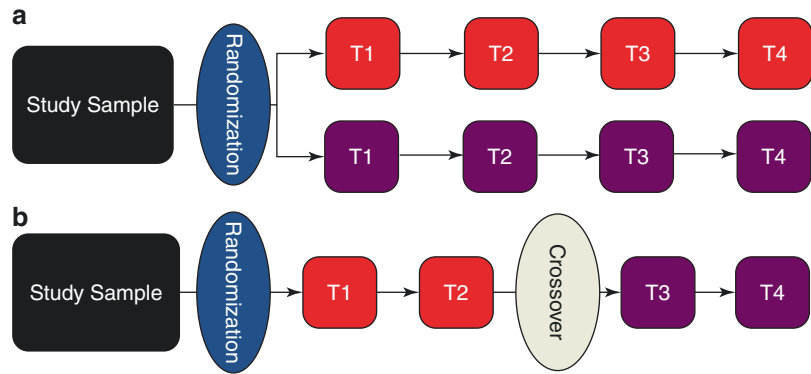
17.1.3 Experimental Research Designs

In comparison to observational studies, *experimental studies* involve the allocation of treatment or intervention to a group of subjects in order to test the benefit of a particular treatment. Data for experimental studies are measured prospectively to determine causal relationships. Studies that are experimental, involve human subjects, and incorporate randomization of group allocation are called *clinical trials*.

There are multiple designs for experimental studies. The “gold standard” in regards to study evidence and establishing causal relationships is the *randomized controlled trial* (RCT). An RCT incorporates a sample of subjects and then assigns individuals to an intervention and control group through unbiased or random allocation.

A type of experimental study design is the *crossover design*. In this study, subjects receive either interventions (or the intervention and control treatment) in a particular sequence (Fig. 17.2a). Due to every subject receiving both interventions, fewer subjects are needed for a crossover study. With this design, it is important to make sure that a washout period is implemented to make sure that the effects of one treatment is not carried into the collection period of the other. In a *parallel design*, independent groups of patients are assigned to only receive one treatment (Fig. 17.2b). Treatment allocation can occur randomly or nonrandomly.

Fig. 17.2 Designs for prospective, experimental research. (a) Parallel study design. (b) Crossover study design



Fact Box 17.1

Level of evidence is based on the quality of the research design in order to influence clinical practice. Less potential threats to validity is represented in higher levels of evidence, resulting in greater cause and effect establishment.

17.2 Hypothesis Testing

A study *population* includes all persons having a common characteristic of clinical or scientific interest. Due to it being impossible to observe or study the entire population, a smaller *sample* of subjects is collected. An assumption is made that our study sample accurately represents the entire population of interest. *Random sampling* rids of bias and allows all members of the population an equal chance to be included in the study sample.

All data collected from a study subject can be categorized as continuous or categorical. *Continuous data* have an infinite number of possible values. Examples of continuous variables are height, weight, age, and time. Conversely *categorical* or *discrete data* have a limited number of possible values. Categorical data could be dichotomous or binary in nature (e.g., failure/success) or categorical (e.g., mild, moderate, severe). Data can be expressed in frequency distributions to summarize or describe characteristics of the study sample.

Descriptive statistics such as mean, median, and mode are measures of *central tendency* used to describe distributions of continuous data. The dispersion of the distribution can be characterized through the standard deviation, range, and percentiles.

The mean is an average calculated as the sum of all scores in the sample divided by the number of subjects in the sample. The median is calculated by dividing a distribution of scores into two equal parts so that half of the scores fall below the median and half of the scores fall above the median. The mode is defined as the value in the distribution that occurs most frequently.

The standard deviation describes the spread, or variability, of the data in a study sample. A higher standard deviation means a higher variability or a wider range of data points. The range is defined as the difference between the largest and smallest observations in the study sample. The spread of the distribution that relate to the rank order of the values can also be characterized by percentiles or sometimes called quantiles. For example, a value x at the 80% percentile of the distribution indicates that 80% of the data points in the sample are less to or equal to x . The median of the sample would be the 50th percentile.

The measures of central tendency and variability can be used to characterize the data frequency distribution. In a *normal* or *Gaussian distribution*, the data set clusters evenly and symmetrically around a value that is the mean, median, and mode. In a normal distribution, 68% of the values are within one standard deviation, 95% of the values are within two standard devia-

tions, and 99.7% of the values are within three standard deviations of the mean.

Skewness is a measure of symmetry or central tendency of the data distribution. Data distributions can be skewed to the left (negative) or to the right (positive). An excessive outlier in the data can often cause skewness of the distribution (Fig. 17.3).

Kurtosis describes the peak or variance of the distribution. A high kurtosis value would be indicative of a high peak in the data distribution or a low sample variance. A low kurtosis value would represent a flat peak in the data distribution or a high sample variance.

17.2.1 Inferential Statistics

Inferential statistical tests are aimed to generalize data collected from a representative sample to the entire population. These statistics can be used to test hypotheses in terms of relationships

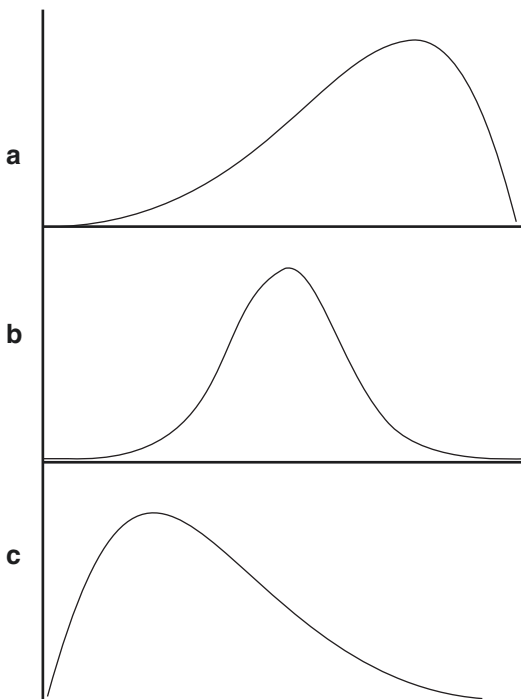


Fig. 17.3 Skewness of a data distribution. (a) Negative (left) skew. (b) Normal distribution. (c) Positive (right) skew

or differences among groupings of data. Inferential statistical tests are divided into parametric and nonparametric statistics. Optimally, researchers want inferences of the study sample to represent a parameter (i.e., the population you are studying), so default statistical tests should be parametric. However, parametric statistics are most powerful when the dataset is normally distributed. If the dataset is skewed or kurtotic, statistical tests are less robust. There are equivalent nonparametric tests that are similar in concept to parametric tests, only more appropriate for datasets that don't meet the assumptions necessary to justify parametric tests. The decision on which test to use is based on the type of data in the data set, the distribution of data, as well as the research question and study design.

Parametric statistical tests use the mean and standard deviation of the distribution to compare groups or identify relationships between variables. *Nonparametric statistical tests* use the medians and ranks of the data and are less sensitive to outliers and more robust. Nonparametric tests are also applied for categorical data and samples with a small sample size.

17.2.2 Tests to Compare Two Groups of Continuous Data

The *t*-test is used to compare continuous variables between two groups and can be used for both paired and independent samples. The *independent sample t-test* (also known as student's *t*-test) is used to compare continuous data collected from two different groups. A between-subject design that aims to compare two groups of data would use the independent sample *t*-test. The nonparametric equivalent of the independent sample *t*-test is the *Mann-Whitney U-test*.

The *paired sample t-test* or the *dependent sample t-test* is used to compare data from within-factor research designs that aims to collect two groups of data—typically from repeated/serial measurements from the same source/subject. The nonparametric equivalent of the paired sample *t*-test is the Wilcoxon signed-rank test.

Clinical Vignette 1

The study *Comparison of Patellar Tendon and Hamstring Tendon Anterior Cruciate Ligament Reconstruction: A 15-Year Follow-Up of a Randomized Controlled Trial* [1] utilized student's *t*-test to compare measures of subjective function, effusion, range of motion, and radiographs in individuals who received a hamstring graft or patellar tendon graft. In this study, the independent variable was the group (hamstring graft, patellar tendon graft), and the dependent variables were clinical outcomes (subjective function, effusion, range of motion, and radiographic outcomes).

17.2.3 Tests to Compare Three or More Independent Groups

The *analysis of variance* (ANOVA) is the statistical test when comparing continuous variables between three or more groups. The independent variable is the nominal or ordered variable used to categorize the data (example groups for a study evaluating knee function: ACL-reconstructed, ACL-deficit, healthy control). The dependent variable is the continuous measure that is the result of manipulating the independent variable (e.g., steps per day). Comparison to a *t*-test that uses the *t*-distribution to compare groups, the ANOVA utilizes the *F*-distribution. The ANOVA will provide an output to inform if the groups are significantly different ($p < 0.05$) or not ($p > 0.05$). The test will not be able to inform where the specific differences within the groups lie. Post hoc (Latin for “after the fact”) tests are then used after obtaining a statistically different ($p < 0.05$) ANOVA to determine exact group differences. Post hoc testing is required as each group comparison may not be the same in a statistically significant ANOVA.

Repeated measures ANOVA are used when the same dependent variables are collected at serial time points. A repeated measures ANOVA assesses the outcome measures in the same subjects in a longitudinal design, assessing the

impact of time. These serial time points could be used to assess effect of an intervention pre- and posttreatment or the progression of disease or illness over time. The *Friedman test* is the nonparametric test equivalent to the repeated measures ANOVA.

When there is more than one dependent variable being compared between groups, a *multivariate analysis of variance* (MANOVA) should be used. A MANOVA may be justified if more than one dependent variable is needed to describe the outcome of interest. For example,

if the outcome of interest is knee degeneration in patients following different surgical techniques, the researchers may want to quantify degeneration in more than one variable (joint space narrowing and the number of osteophytes). For this example, the different surgical techniques would be the independent variable, and the joint space and number of osteophytes observed from radiographic imaging would be our dependent variables.

An *analysis of covariance* (ANCOVA) is a statistical test used when a confounding variable needs to be accounted for. The ANOVA then can be adjusted to establish groups' differences using the covariate as a “statistical control.”

Similar to the ANOVA, post hoc testing is needed to determine group differences for repeated measures, ANCOVA, and MANOVA testing. The *Kruskal-Wallis test* is the nonparametric test equivalent to the ANOVA and should be used when the data is categorical or not normally distributed and samples are independent. Similar to Wilcoxon rank-sum test, the Kruskal-Wallis pools the observations from all comparison groups and assigns ranks to each. The average ranks are then compared between groups rather than group means. Difference among the average ranks determines whether there are differences among groups. Much like the ANOVA models, a nonparametric post hoc test would be required to determine the exact differences that exist among multiple (three or more) groups of data.

In addition to assessing the differences of more than two groups, an ANOVA can be used to assess effects and interactions when multiple independent variables are of interest. Common

grouping variables in research designs are sex (female, male) and treatment (treatment, control). This would be analyzed through a 2×2 factorial design due to both independent variables having two levels. Factorial designs that are normally distributed can be analyzed through ANOVAs, MANOVAs, ANCOVAs, or a repeated measures ANOVA (often called a split-plot design). A factorial repeated measures ANOVA (or split-plot ANOVA) compares the multiple groups over serial time points. For example, a research study which aimed to see the effect of an intra-articular injection (treatment group, corticosteroid; control group, saline) on knee joint pain over time (1-week postinjection, 2-week postinjection, 3-week postinjection) would utilize this study design. The two factors in this study would be group and time. This design would allow the ability to see the difference in pain between groups, time points, and/or an interaction of group (i.e., treatment allocation) and time.

Clinical Vignette 2

The study *Age Influences Biomechanical Changes after Participation in an Anterior Cruciate Ligament Injury Prevention Program* utilized a Univariate ANOVA to assess biomechanical differences between preadolescents and adolescent athletes who did and did not participate in a training program [2].

other set alpha) is observed, the test would be deemed statistically significant. Therefore, the conclusion would be to reject the null hypothesis and that there are statistical differences between groups. If the *p-value* is above the established alpha, the results conclude that any observed differences would be due to chance and not due to any interventions or grouping, therefore, accepting the null hypothesis.

A Bonferroni correction is used to adjust the *p-value* or level needed for statistical significance, when more than one comparison is being performed between the groups (Table 17.1). As the number of comparisons you make between the group increases, your likelihood of finding a statistical significant result also increases. To prevent this, or commonly referred to as “protecting your alpha,” a Bonferroni correction will be applied to lower your threshold of significance for a greater number of comparisons being made. This is applied by dividing your alpha (typically 0.05) by the number of comparisons you are making. This new threshold for defining statistical significance is an adjustment due to the possibility of error from making multiple comparisons.

Statistical power is the ability to detect group differences or an association when one actually exists. *Type II errors (beta)* occurs when a statistical test infers that there are no differences between groups when one actually exists. A typical established *beta* is 0.2 or up to 20% of the time. Therefore, statistical power ($1 - \beta$) should be greater or equal to 0.80 or able to detect differences or relationships between variables at least

17.2.4 Determination of Significance

Probability values (p-value) are associated with inferential statistics to determine if a test statistic is statistically significant. Traditionally, a *p-value* (or *alpha*) is set at a threshold of 0.05. This *p-value* refers to the probability that the result from the statistical test is due to purely chance alone or, in other words, would occur once out of 20 tests. The null hypothesis for any statistical comparison is that there are no differences between groups. If a *p-value* less than 0.05 (or

Table 17.1 Interpretation of the *p-value*

Null hypothesis	$H_0 = H_1$	<i>p-value</i> > a priori set <i>alpha</i>	Accept the null hypothesis. Conclude no differences between groups
Alternate hypothesis	$H_0 \neq H_1$	<i>p-value</i> < a priori set <i>alpha</i>	Reject the null hypothesis. Conclude differences between groups

Fact Box 17.2

“Protecting Alpha Error” If you are going to make three comparisons between your study groups, a Bonferroni correction can be used as follows: $0.05 \div 3 = 0.017$. The alpha level of a test would need to be ≤ 0.017 (compared to 0.05) to be considered statistically significant.

80% of the time if they exist. Statistical power is increased when sample sizes increase or when data variability decreases. *Power analyses* utilize these relationships and can be performed from established surrogate data or preliminary pilot data to determine how many subjects are required to provide the greatest chance of finding a statistically significant result (Table 17.2). Properly preformed power analyses should be performed with limiting the possibility of type I error to 5% ($p = 0.05$) and with statistical power being greater than 80% (type II error).

A *confidence interval* quantifies the variance error around the point estimation of the study sample. Confidence intervals are calculated from the study means and measures of variance of the distribution are which a population parameter will fall. Samples with a greater variability will have larger confidence intervals. A 95% confidence interval is often utilized in clinical research and describes a range of values in which researchers are 95% certain that the actual measure of the population will fall within those values.

Effect sizes measure the magnitude of the treatment effect and are often useful when determining the clinical importance of research find-

ings. A stronger effect size will be present with greater mean differences and smaller data variability. Effect sizes range from zero to infinity, with values closer to zero representing a weaker effect. One example of effect size classification is Cohen’s D. Cohen’s D effect size values are considered “weak” if less than 0.2, “medium” if less than 0.5, and “strong” if greater than 0.8.

17.2.5 Tests for Categorical Data

The *chi-square* (χ^2) test is an appropriate test when there are two or more groups of categorical data. Rather than the distribution used in continuous data, the frequency of results would be analyzed using the chi-square test. *Fisher’s exact* test is another test used to analyze categorical data; however, it is more so utilized for small samples or when one or more categories contain few or no data points.

Fact Box 17.3

Inferential statistics are used to test specific hypotheses of study subjects or sample a come measure of interest. The independent variables are the groupings of the data that is observed or manipulated by the investigator.

17.2.6 Measures of Association

Correlations are used to describe the strength of the relationship or association between two variables. The *Pearson product-moment correlation coefficient* (r) is the test utilized to assess the relationship between two normally distributed, continuous variables. The r value can range from -1 to 0 to $+1$. Values closer to -1 or $+1$ represent stronger relationships and values closer to 0 represent weaker relationships. Positive values represent directive relationships; a high value in one variable would often be seen with a high value in the other variable. Conversely, a negative value would represent an indirective relationship. A

Table 17.2 Factors affecting sample size

Condition	Effect on sample size
Variability in outcome measure increases	Increases
Significance level (alpha) decreases	Increases
Required power ($1 - \beta$) increases	Increases
Effect size or mean difference increases between groups	Decreases

negative r value would be indicative of a high value in one variable associating to a low value of the other variable. Pearson classification considers correlation coefficients less than 0.33 “weak,” those less than 0.66 “moderate,” and those greater than 0.66 “strong.”

The nonparametric equivalent to the Pearson correlations coefficient is *Spearman’s rho*, which should be used for non-normally distributed data or categorical data.

Regression describes the ability to predict a specific outcome variable and is expressed in a *coefficient of determination* (R^2). In comparison to correlation, a regression analysis has an intended outcome variable that is explained by one or more predictor variables. An outcome variable is the variable singled out to be predicted by the other variables, often called the dependent variable. Predictor variables, or independent variables, are variables used to make the prediction. The use of one predictor is known as *simple linear regression* compared to the utility of multiple predictors, known as *multiple regression*. The coefficient of determination ranges from 0 to 1, where a higher value is indicative of a greater proportion (%) of variance explained and a better predictive value. Both simple linear and multiple regression are used when the outcome variable is continuous. When the outcome variable is categorical (often dichotomous) or not normally distributed, *logistic regression* should be used.

17.2.7 Tests of Agreement Between Variables (ICC and Kappa)

The *intraclass correlation* (ICC) should be used to assess the consistency or reproducibility of quantitative measures. These statistics are similar to other correlation coefficients, only they assess agreement between arrays of data. These tests are often used in reliability and validity research studies to evaluate concordance among two outcomes or measurements of interest. An intraclass correlation could also be used to assess the same outcome being examined by different individuals. For example, a radiograph of knee osteoarthritis could be sent to multiple reading evaluators

Table 17.3 Interpretation of the *intraclass correlation coefficient* and *Cohen’s kappa coefficient*

Intraclass correlation (ρ_i)		Cohen’s kappa (κ)
Poor reproducibility	<0.4	Marginal reproducibility
Fair to good reproducibility	≥0.4, <0.75	Good reproducibility
Excellent reproducibility	≥0.75	Excellent reproducibility

to determine the consistency of the results. Interpretation of the *intraclass correlation coefficient* (ICC or ρ_i) is defined in Table 17.3.

Cohen’s kappa coefficient (κ) is a measure of interrater agreement for categorical data. Kappa is usually used as a measure of reproducibility of

Clinical Vignette 3

In the study *The Reliability of Assessing Radiographic Healing of Osteochondritis Dissecans of the Knee*, ICC values were used to determine interrater and intrarater reliability of assessing OCD healing on plain radiographs [3]. In this study, multiple surgeons evaluated the radiographic healing of the knee at two time points, minimally 1 month apart. Healing of OCD lesions was found to have excellent interrater reliability (ICC = 0.94), indicating high agreement among the raters.

Fact Box 17.4

Statistical tests for comparisons should match the purpose of the pursued research study. Analyses should be run to accomplish the goals of the researcher. If the interest of the study is to find differences between groups, t -tests or ANOVA’s should be performed. If the goal of the study is to find associations or prediction of variables, correlations or regression analyses should be performed.

qualitative outcomes between repeated assessments of the same variable. Guidelines for evaluating kappa are in Table 17.3.

Fact Box 17.5

- Level of evidence is based on the quality of the research design in order to influence clinical practice. Less potential threats to validity is represented in higher levels of evidence, resulting in greater cause and effect establishment (Fig. 17.2).
- Inferential statistics are used to test specific hypotheses of study subjects or sample data. The dependent variable is the outcome measure of interest. The independent variables are the groupings of the data that is observed or manipulated by the investigator.
- Statistical tests for comparisons should match the purpose of the pursued research study. Analyses should be run to accomplish the goals of the researcher. If the interest of the study is to find differences between groups, *t*-tests or ANOVA's should be performed. IF the goal of the study is to find associations or prediction of variables, correlations or regression analyses should be performed.

Take Home Messages

- The first step is selecting a statistical analysis is to determine they type of level of the data that will be analyzed.
- Statistical significance tells us that the finding unlikely occurred by chance, rather that it is a reflection of the population. Statistical signifi-

cance does not always mean clinical meaningfulness.

17.3 Useful Resources

Hung M, Bounsanga J, Voss MW. Interpretation of correlations in clinical research. *Postgraduate Medicine*. 2017;129(8):902–906 [4].

Barkan H. Statistics in clinical research: Important considerations. *Ann Card Anaesth*. 2015;18(1):74–82 [5].

Kocher MS, Zurakowski D. Clinical epidemiology and biostatistics: A primer for orthopaedic surgeons. *Journal of Bone and Joint Surgery-American Volume*. 2004;86A(3):607–620 [6].

Petrie A. Statistics in orthopaedic papers. *Journal of Bone and Joint Surgery-British Volume*. 2006;88B(9):1121–1136 [7].

References

1. Webster KE, Feller JA, Hartnett N, Leigh WB, Richmond AK. Comparison of patellar tendon and hamstring tendon anterior cruciate ligament reconstruction: a 15-year follow-up of a randomized controlled trial. *Am J Sports Med*. 2016;44(1):83–90.
2. Thompson-Kolesar JA, Gatewood CT, Tran AA, Silder A, Shultz R, Delp SL, et al. Age influences biomechanical changes after participation in an anterior cruciate ligament injury prevention program. *Am J Sports Med*. 2018;46(3):598–606. <https://doi.org/10.1177/0363546517744313>.
3. Wall EJ, Milewski MD, Carey JL, Shea KG, Ganley TJ, Polousky JD, et al. The reliability of assessing radiographic healing of osteochondritis dissecans of the knee. *Am J Sports Med*. 2017;45(6):1370–5.
4. Hung M, Bounsanga J, Voss MW. Interpretation of correlations in clinical research. *Postgrad Med*. 2017;129(8):902–6.
5. Barkan H. Statistics in clinical research: important considerations. *Ann Card Anaesth*. 2015;18(1):74–82.
6. Kocher MS, Zurakowski D. Clinical epidemiology and biostatistics: a primer for orthopaedic surgeons. *J Bone Joint Surg Am*. 2004;86A(3):607–20.
7. Petrie A. Statistics in orthopaedic papers. *J Bone Joint Surg Br*. 2006;88B(9):1121–36.

Clair Smith

18.1 List of Definitions

Data: A collection of data points organized into one or more variables of interest.

Example: The set of all responses to a survey given to a group of people, all measurements taken from the mice in an animal study, etc.

Variable: A measureable characteristic such as blood pressure, age, or gender.

Example: Treatment group, marital status, diabetes status, systolic blood pressure, blood glucose level, etc.

Observation: A single datum. In clinical data this will often be a measurement taken from a person or animal. Also called a data element or data point.

Example: The heart rate of mouse in an animal study, the cancer status of a cell from a person in a cancer study, the range of motion of a knee from a cadaver in a meniscectomy study, and the BMI of one person in a study.

Statistic: A numerical summary of the data points that make up a variable. This can be calculated from a sample.

Example: Mean, variance, median, minimum, and maximum.

Sample: Data that is collected/observed. A small subset of the population of interest is presented.

Example: A random sample of residents of a certain neighborhood and all people hospitalized for a heart attack at one of three local hospitals during a certain period of time.

Population: The group of all subjects researchers are interested in studying.

Example: The set of all women currently living in the USA, the set of all people experiencing lower back pain in the USA, and the set of all mice of a certain species.

18.2 Types of Data

There are two major types of data that researchers typically deal with in health science: continuous and discrete data. The type of data drives which statistics are used in their analyses. Continuous data such as age, height, weight, and BMI have infinitely many possible values. For example, age in years can be any positive real number such as 42 or 37.25. Discrete (also known as categorical) data has a limited number of values it can take on such as race, treatment group, and study site. For example, if possible values of race on a self-reported survey are black, white, and other, then everyone taking the survey will have one of these three values for the variable

C. Smith (✉)
Department of Orthopaedic Surgery,
University of Pittsburgh, Pittsburgh, PA, USA

Department of Physical Therapy,
University of Pittsburgh, Pittsburgh, PA, USA
Bridgeside Point 1, Pittsburgh, PA, USA
e-mail: cns45@pitt.edu

race. Other less common types of data are count data and censored data. Count data is made up of whole numbers that represent counts such as the number of falls in a year of follow-up or the number of heartbeats per minute. While there are analyses created specifically for count data, it is often treated as continuous data for simplicity. Censored data, such as number of years till death after having a certain procedure, occurs when the event researchers are interested in (such as death) may not occur during the study. Another type of data, longitudinal data, occurs when measurements are taken repeatedly from subjects over a period of time.

A continuous variable is one whose values can be any real number. It is meaningful to measure the distance between values, and arithmetic operations such as addition and multiplication make sense for continuous variables but not for discrete variables. Technically, all continuous variables are measured discretely since we don't have instruments that can be measured continuously. One can think of continuous data as discrete with lots of levels or categories. For example, blood pressure is typically measured to 2 mmHg because measuring with higher precision would be difficult with the instruments that are used. However, we still treat blood pressure as continuous since there would be far too many levels to treat it as discrete. Sometimes discrete variables with many levels such as the visual analog scale (VAS) for measuring pain or variables measured with the Likert scale are treated as continuous for ease of analysis. When treating discrete variables as continuous, you are assuming that each level of the variable is equidistant apart. Another example of a continuous variable is the Lysholm scale for assessing ACL injuries which gives a score from 0 to 100 with higher scores indicating fewer symptoms.

There are two major types of discrete data: nominal and ordinal. Nominal data has no inherent ordering such as gender, race, and marital status. Ordinal data can be ordered from low to high such as injury severity, level of education, and household income. The differences between the levels of ordinal variables are not necessarily equal. For example, the difference

between the mild and moderate level of an injury severity variable may not be the same as the difference between the moderate and severe level. For both types of discrete data, each observation must belong to exactly one level of the discrete variable, and the levels should cover all possible values that exist in the data set. For example, if the discrete variable race has levels black, white, and other, then each observation in the data set must be categorized as black, white, or other. If a discrete variable has only two levels, then it is called a dichotomous variable. Examples include gender and disease status (the disease is either present or not present).

18.3 Data Description

Summarizing discrete data is simpler than summarizing continuous data. Discrete data is often described by reporting the frequency and proportion (or percent) of people belonging to each level of the discrete variable. For example, say you were reporting on disease severity in a study of 50 people. Your description of the variable "disease severity" could be 23 (46%) mild, 12 (24%) moderate, and 15 (30%) severe if 23 people in the study had mild disease, 12 had moderate, and 15 had severe. If the variable is dichotomous, then it is acceptable to report only the frequency and proportion in one level of that variable. For example, if researchers were summarizing the dichotomous variable "gender" and putting it in a table of demographic information for a study, then they could simply report the frequency and proportion of women in the study sample. Researchers would not need to include the number of men in the study since this can be deduced by subtracting the number of women in the study from the sample size. Some researchers depict the proportions of subjects in each level of a discrete variable in a bar graph. This may be appropriate if the publication has no other figures. However, when other figures are present, graphing proportions is superfluous as they are described adequately by frequencies and proportions alone.

The relationship between two categorical variables is best captured by a 2×2 table. In such a table, the rows are levels of one categorical variable, and the columns are levels of the other categorical variable. The cells of the table contain the number of people in the study belonging to the corresponding levels of the row and column variables. The last row and last column are typically reserved for totals (also known as margins).

Continuous data contains more information, has more properties, and requires more statistics to describe it than discrete data. There are different statistics to measure the location (or center), spread (or dispersion), and shape of the distribution of values from a continuous variable.

Measures of location seek to describe the central tendency of the data with a representative value from it. Examples are the mean (or average), median, and mode. The mean of a continuous variable is the sum of all the values divided by the number of values present in that sum. Put

into symbols the mean is $\frac{\sum_{i=1}^n x_i}{n}$, where x_i repre-

sents the i th value, n is the sample size, and $\sum_{i=1}^n x_i$ says to sum all the values from 1 to n . The letter i is called an index. The median is the middle value of the ordered data. If the values are ordered from smallest to largest (or largest to smallest), then the median is the value that has an equal number of observations on either side of it. If the sample size is even, then the median is found by averaging the two middle values of the ordered data. Instead of listing out the ordered values by hand and visually finding the middle value, there is a simple formula that can be used to find the position of the median. Say there are n people in the study and the age of each person is listed from smallest to largest. If n is odd, then the position of

the median age is $\frac{n+1}{2}$. Note that this equation will produce the *position* of the median, not the value of the median itself. If n is even, then the

median is the average of the numbers in the $\frac{n}{2}$

and $\frac{n}{2} + 1$ positions of the ordered list of values

[1]. The median and mean can only be used to describe continuous data. The mode of a variable is the most frequently occurring value and can be used to describe the central tendency of continuous or discrete data.

Another name for the median is the second quartile. The quartiles split the list of ordered values into fourths. The first quartile is also called the 25th percentile and is often denoted Q_1 . If the data is listed in order, then 25% of the values will be below or equal to the first quartile. The median is the 50th percentile (or second quartile) since half of the values are less than or equal to it and it is denoted Q_2 . The third quartile, Q_3 , is also called the 75th percentile, and 75% of the values are less than or equal to it.

Measures of spread describe how tightly clustered the values are around the mean of continuous data. Examples are the variance, standard deviation, range, and interquartile range. The variance is the average squared distance from the mean. It is calculated by adding up all of the squared differences between the mean and each data point then dividing this sum by the number of data points minus one. If n is the sample size and \bar{x} is the mean of the sample, then the following is an equation for finding the variance of the

sample: $\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$. Note that Σ is a symbol

meaning “the sum of” and x_i represents the i th value in the sample. When put together as in

$\sum_{i=1}^n (x_i - \bar{x})^2$, this means take each value from the

first ($i = 1$) to the last ($i = n$), subtract the mean from it, then square it, then add all these squares together. This equation is similar to the equation for the mean except that it is divided by $n - 1$ instead of n . Dividing by $n - 1$ leads to a less biased statistic than dividing by n . The standard deviation is the square root of the variance and

thus the average distance from the mean. The range is simply the maximum (largest) value minus the minimum (smallest) value. The interquartile range (IQR) is the third quartile minus the first quartile. The IQR describes the spread of the central 50% of the data. For all measures of spread, a higher value indicates that observations are more spread around the mean and smaller values indicate they are more tightly clustered about the mean.

Fact Box 18.1

$$\text{Mean} = \frac{\sum_{i=1}^n x_i}{n}$$

$$\text{Position of median for odd } n = \frac{n+1}{2}$$

$$\text{Position of median for even } n = \frac{\frac{n}{2} + \left(\frac{n}{2} + 1\right)}{2}$$

Measures of the shape of a distribution describe the overall trend of the data. As an example, a distribution could have mostly large values with a few extreme outliers, or it could have values evenly distributed across the range. Some examples of distribution shapes are symmetric, normal, bimodal, and left or right skewed. The mean, median, and mode of a continuous variable are equal if the distribution of its values is symmetric. In terms of symmetric data, the relative position of observations is the same on either side of the median. Right skewed (or positively skewed) data occurs when observations above the median are farther in absolute value than observations below the median. Another way of saying this is that the distribution has a long tail to the right. In right skewed data, the majority of the values are relatively small and close together, and a minority of the values are extreme or much larger in value than the rest. Left skewed (or negatively skewed) data occurs when observations below the median

are farther in absolute value than observations above the median. Left skewed data has a long tail to the left since most of the values are large and there are a few extreme observations that are much smaller than the rest. The mean is greater than the median in right skewed data and less than the median in left skewed data. There is a skewness index that measures the degree of skewness in the data. The index is zero if the data is symmetric, greater than zero if the data is right skewed, and less than zero if the data is left skewed [3]. A *normal distribution* is a symmetrical hill or bell shape with the majority of the values close to the central value (the mean) and a few extreme observations on either side of the mean (i.e., in the tails of the distribution). A bimodal distribution looks like the two humps of a camel; it has two central values. When the data is symmetric, the best numerical summaries are the mean and standard deviation. When the data is skewed, it is best to use the median and interquartile range or interquartile deviation (half of the interquartile range).

Fact Box 18.2

$$\text{Variance} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

$$\text{Standard deviation} = \sqrt{\text{variance}}$$

$$\text{IQR} = Q_3 - Q_1$$

Kurtosis is another measure of shape that describes how flat or steep the distribution of values is compared to a bell-shaped (or *normal*) distribution. If there are a lot more observations in the tails of a distribution compared with a normal distribution, then the graph appears flatter than a bell shape. If there are many fewer observations in the tails of a distribution compared with a normal distribution, then the graph appears more peaked than a bell shape [2].

All of the descriptive measures discussed thus far are statistics. Statistics are calculated from a sample that is drawn from the population of interest. Suppose the goal of a study is to determine whether a new surgical technique for repairing a joint leads to a better clinical outcome than the standard procedure. The population of interest in this case would be the set of all people with the joint injury who would be eligible for this surgery. In order to determine whether the new technique is an improvement over the old technique, researchers must look at the outcomes from a sample of people with the joint injury. It is not possible to observe all people with this injury (the population of interest), so a sample must be taken. Typically, the sample is chosen in such a way that every member of the population has an equal opportunity of being picked for the sample. A sample that is created in this way is called a random sample because each member is chosen at random. This helps to ensure that the sample is representative of the population, e.g., if half of the population is women, then roughly half of the random sample drawn from the population should be women. The researchers would then use statistics such as means and standard deviations calculated from this sample to summarize outcomes of the two surgery techniques. Such an outcome may be the range of motion of the repaired joint after it has healed. Since range of motion is a continuous measure, researchers would use a mean or median and standard deviation or interquartile range to summarize it. This example illustrates using a sample to make inference about a population, the main goal of statistics.

Since a sample does not include all members of a population, there are multiple ways to draw a sample from a given population. The number of people in the population of interest is typically denoted by N , and the number of people in a given sample drawn from the population is denoted by n . Figure 18.1 depicts three different samples of size n drawn from a population of size N .

Suppose the three samples in the figure were drawn in a sequence: n individuals were selected at random from the population to

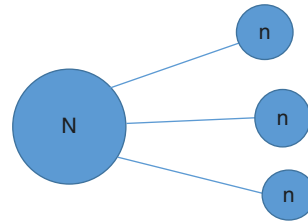


Fig. 18.1 Samples of size n drawn from a population of size N

form the first sample, measurements were taken on them, and then they were returned to the population pool. This way of sampling is called *sampling with replacement*. This process was then repeated for the second and third samples of size n . If we calculated the mean of the measurements taken on each of the three samples, we would get three different means even though the samples are the same size and are taken from the same population. This occurs because the three samples consist of different individuals. It is possible that there is overlap between the three samples, that is, some individuals may occur in two or more of them. This is possible because the samples were drawn from the population with replacement: they were selected, their measurements taken, and then they were returned to the population pool. Thus, every time a sample is drawn from the population, a different sample mean is calculated, but each of these means will be a good estimate of the true mean of the entire population (given that the sample size, n , is sufficiently large). The mean calculated from the sample of size n is an example of a sample statistic, and the mean calculated from the population of size N is an example of a population parameter. Sample statistics are estimates of population parameters. Population parameters are usually unknown since we cannot measure an entire population but we estimate these parameters by taking a random sample of the population and calculating sample statistics. The larger the sample size, the more confident researchers are that the sample statistics are good approximations of the population parameters.

18.4 Visual Displays

It is good practice to plot the data before summarizing and performing statistical tests on it. This will give the researcher a sense of the type of data available. There are various methods for describing and analyzing data, and which method to be used depends on the nature of the data.

A stem-and-leaf plot is a simple visual display of data points that shows the distribution or shape of the values of a continuous variable. An advantage of this plot is that it includes the value of each individual observation. This plot is appropriate when there are a small number of observations. As an example, suppose there is a small sample of 15 subject's BMI measurements that have been rounded to the nearest integer. BMI is a continuous variable and its units are kg/m^2 . The first step of making a stem-and-leaf plot of these values would be to list them in order:

18,19,23,24,24,24,25,25,26,26,27,28,30,32,37

The stem of the plot is made up of the leading number, and the leaves are made up of the trailing number. Both the numbers in the stem and in the leaves are ordered smallest to largest.

```

1 | 89
2 | 3444556678
3 | 027

```

It can be seen from this plot that the BMI measurements are distributed in roughly a hill shape: most of the values are in the middle and there are a few in either tail. If there are more observations, more than one line can be added for each digit in the stem.

A histogram is a graph that shows the shape of the distribution of values of a continuous variable. The horizontal (or x) axis has the values of the variable, and the vertical (or y) axis has the frequency or proportion of observations. The height of each rectangle represents the proportion or frequency of observations whose values fall within the range specified by the width of the rectangle. If the distribution of values of a variable is symmetric, then cutting the histogram

along the median will result in each half being a mirror image of the other. A common example of a symmetric distribution is a hill or bell-shaped distribution. Figure 18.2 shows a histogram for the 15 BMI measurements in the last example.

The width of the rectangles in this histogram is five observations, and the vertical axis is the frequency of occurrences. The x -axis shows the range of values for each rectangle. The histogram shows the general *hill-shaped* trend of the data: most of the data (9 observations) fall within the range of 23–28 kg/m^2 . This histogram shows a similar shape as the stem-and-leaf plot turned on its side. If the width of the rectangles is too large, important information about the shape of the distribution can be lost. The smaller the width of the rectangles, the more detail about the shape of the distribution will be shown. Most statistical programs will automatically choose a width that is appropriate for the data.

Another visual display for continuous data is the box plot. The box plot shows the interquartile range, the median, and any extreme observations (i.e., observations that have values that are much larger or much smaller than the rest of the data). If there is a lot of variability in the data, then the box and whiskers will be elongated. If there is not a lot of variability, then the box and whiskers will appear squatter. Figure 18.3 shows a box plot of the BMI example data.

The first quartile of the BMI data is the bottom line of the box, the median is the middle line in the box, and the third quartile is the top line of the box. When the third quartile is farther from the median than the first quartile, the data is *right skewed*, and

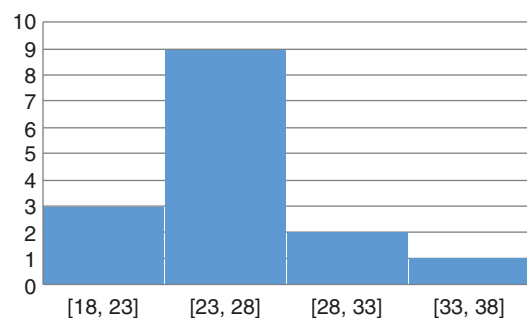


Fig. 18.2 Histogram of a sample of 15 BMI measurements

when the first quartile is farther from the median, the data is *left skewed*. The histogram and box plot of the BMI data in Fig. 18.3 show a slight right skew in the shape of the distribution. The whiskers of the box plot are drawn to the smallest and largest observations in the sample that are not outliers. Outliers are defined as values that are greater than $Q_3 + 1.5*(IQR)$ or less than $Q_1 - 1.5*(IQR)$. The dots in the box plot are extreme outliers, which are defined to be larger than $Q_3 + 3*(IQR)$ or smaller than $Q_1 - 3*(IQR)$ [3].

The box plot is a good visual display to use when comparing a continuous variable between different groups of a categorical variable since they can be plotted side-by-side on the same set

of axes. This allows direct comparison of the distributions of the continuous variable across various levels of the categorical variable. As an example, consider the BMI data again, but suppose there is information on whether the subjects were over 25 years old or under 25 years old. Figure 18.4 is an example of a way to visualize the relationship between a continuous variable (BMI) and a categorical variable (age category).

From Fig. 18.4 it can be seen that subjects who are over 25 years old have a higher BMI than people who are 25 years old or younger.

A scatter plot is useful for understanding the relationship between two continuous variables and revealing potential outliers. Values of one variable are plotted on the horizontal axis, and values of the other variable are on the vertical axis. A scatter plot is a quick way to discover potential trends in the data. For example, if higher values of one variable tend to occur with higher values of the other, then the scatter plot will show this positive relationship. If there is no relationship between the two variables, then the scatter plot will show a random scattering of points that don't indicate any specific pattern. If most of the points are clustered tightly together, while one or two points are clearly outside of this cluster, then these points are potential outliers and should be checked for accuracy. Figure 18.5 is an example of a scatter plot using the BMI data. A second variable, age, has been added to the vertical axis.

Figure 18.5 shows that as age increases so does BMI. In other words, there is a positive relationship between age and BMI. An example of an outlier for this data is the point (32, 20), i.e., the point with BMI = 32 and age = 20. While

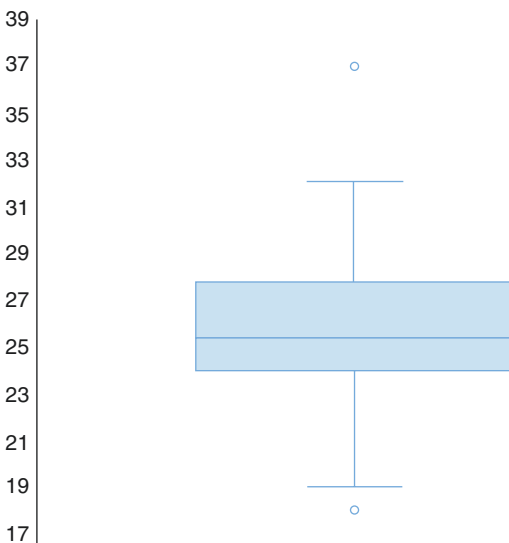


Fig. 18.3 Box plot of 15 BMI measurements

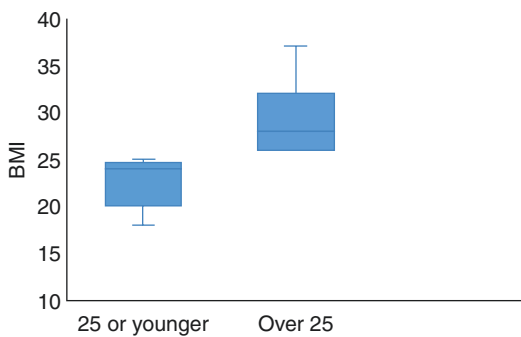


Fig. 18.4 Side-by-side box plots of 15 BMI measurements

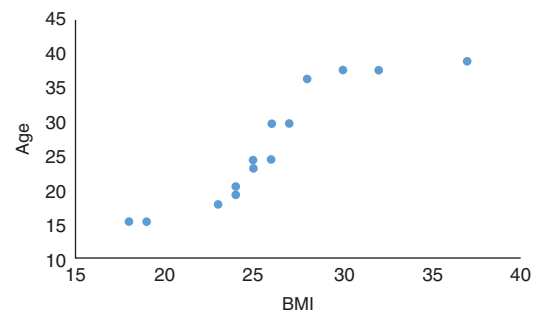


Fig. 18.5 Scatter plot of 15 subjects' age and BMI

the point is medically feasible, if it were in the plot in Fig. 18.5, we would want to check on its accuracy, because it is so far away from the increasing trend of the rest of the points. Two continuous variables could also have a negative relationship if it were the case that as one increased the other decreased. If the scatter plot appears to show a positive relationship for some values and a negative relationship for others, we would say the relationship appears to change direction. A statistic called a correlation coefficient classifies the strength of the association between two continuous variables.

18.5 Conclusion

The first steps of data analysis should be to determine what types of variables are present and to describe them with appropriate summary

statistics and visual displays. Different types of data have different properties, and these properties determine which statistical tests are appropriate for answering the questions of interest. Statistical tests are based on probability theory and allow the researchers to draw conclusions about a population based on a sample from that particular population. This is called statistical inference and is the overarching goal of statistical analysis.

References

1. D'Agostino RB, Sullivan LM, Beiser AS. *Introductory applied biostatistics*. Belmont: Thomson Brooks/Cole; 2006.
2. Daniel WW, Cross CL. *Biostatistics: a foundation for analysis in the health sciences*. 10th ed. Hoboken: Wiley; 2013.
3. Rosner B. *Fundamentals of biostatistics*. 8th ed. Boston: Cengage Learning; 2016.



Does No Difference Really Mean No Difference?

19

Carola F. van Eck, Marcio Bottene Villa Albers,
Andrew J. Sheean, and Freddie H. Fu

19.1 Introduction

In conducting research, it is important to form a sound hypothesis that allows for determination of the proper primary outcome measure for a particular study. Additionally, the hypothesis dictates which statistical methods should be applied. A hypothesis is a proposed explanation for a phenomenon. The null hypothesis (H₀) is the default position and states there is no difference between the test and control group, whereas the one hypothesis (H₁) is its rival and states superiority of one procedure/intervention over another. The *p*-value is a measure of how much evidence there is against the null hypothesis. The null hypothesis is commonly rejected when the *p*-value is less than 0.05 corresponding to a 5% chance of rejecting H₀ when in fact it is true. Type I and II errors

are errors in statistical analysis. A type I (or “alpha”) error refers to the rejection of the H₀ when it is really true. However, the most common statistical error is type II (or “beta”) error: failing to reject a H₀ when in fact it should have been rejected or, more simply said, unjustly concluding there is “no difference” between the variables in question. The purposes of this chapter are to (1) outline several of the more common causes of type II errors and (2) describe strategies for mitigating the likelihood committing type II errors when conducting orthopedic research.

The most common statistical error is type II (or “beta”) error: failing to reject a H₀ when in fact it should have been rejected or, more simply said, unjustly concluding there is “no difference.”

The material in this manuscript has never been published in its current format. The authors did not receive any outside funding or grants directly related to the research presented in this manuscript. The Department of Orthopaedic Surgery from the University of Pittsburgh receives funding from Smith and Nephew, Depuy Mitek, ConMed Linvatec, Cook MyoSite, and Arthrex not directly related to the research presented in this manuscript. Dr. Freddie H. Fu is an editorial board member of AJSM and KSSTA.

C. F. van Eck · M. B. V. Albers · A. J. Sheean
F. H. Fu (✉)
Department of Orthopaedic Surgery, University of
Pittsburgh, Pittsburgh, PA, USA
e-mail: ffu@upmc.edu

19.2 Lessons Learned from Cardiology

Cardiology is perhaps one of the most advanced fields in medicine with regard to research and innovation. The first electrocardiogram (EKG) was performed by Einthoven as early as 1903 [76]. His goal was to develop a simple and inexpensive test which could identify conduction abnormalities of the heart and demonstrate

the anatomic location of the problem [76]. This invention was soon followed by the development of angiography which could not only visualize the location and severity of the problem but also allowed intervention to be performed [6]. More recently, innovation within the field of cardiology catalyzed the integration of high-resolution computed tomography (CT) scanning to more precisely understand the anatomy of the heart's chambers, valves, and blood vessels [54]. This noninvasive test can be performed in 10 s and allows detection of coronary artery disease involving plaques as small as 1.5 mm [77]. This highly specialized application of CT technology has excelled research efforts focusing on treatment of coronary artery disease and other cardiac conditions. However, within orthopedic surgery, this standard has not yet been paralleled.

19.3 Mission Accomplished in Orthopedic Surgery?

Anterior cruciate ligament (ACL) injury is one of the most widely studied topics within the field of orthopedic sports medicine. Some of the earliest described techniques for reconstructing the torn ligament involved a large arthrotomy [42]. However, as with all modern surgery, minimally invasive surgical techniques were introduced in knee surgery, leading to the development of arthroscopically assisted ACL reconstruction [53, 59, 64]. Arthroscopic ACL reconstruction was first performed using a two-incision technique, in which the femoral bone tunnel was drilled from the outside-in [53, 64]. Over time, a one-incision technique was adopted, where the femoral bone tunnel is drilled from inside-out, through the tibial tunnel (transtibial technique) [17]. Both techniques were fast and efficient; unfortunately, neither technique afforded focus on native ACL anatomy [33]. Outcome-based research focused on these minimally invasive techniques relied mostly on subjective patient-reported outcome measures (PRO), which suggested good results.

However, the surgeons implementing these PRO began to observe a variety of postoperative problems, including loss of knee range of motion [16], impingement of the new ACL graft [13, 24, 25, 52], and failure of the reconstruction requiring revision surgery [29]. In mid- to long-term follow-up, a large number of patients were found to have developed osteoarthritis [15]. In this respect, the limitations of early PRO partially obscured an adequate appreciation for some of the complications of ACL reconstruction. In these respects, early outcomes demonstrated that the state of the art in ACL reconstruction left much to be desired.

19.4 Why Our Outcomes Are So Good

Despite surgeons observing the loss of knee range of motion [16], impingement [13, 24, 25, 52], graft failure, and early osteoarthritis [29] aforementioned, orthopedic publications continued to report good results after transtibial ACL reconstruction. The discrepancy between the poor observed outcomes and the good reported outcomes is likely due to the widespread use of instruments that were not truly sensitive to actual outcomes and the flawed interpretations that followed. As recently as 2011, a poll of orthopedic surgeons demonstrated that 83% of respondents based their assessment of surgical outcomes solely on whether the patient was satisfied, rather than on KT arthrometer, pivot shift results, or long-term clinical follow-up [62]. In spite of technically imperfect surgeries, acceptable patient satisfaction scores were observed. However, it is entirely possible that this phenomenon is attributable to placebo effect, which has been shown to exist in approximately 35% of patients. Moreover, it has been shown that 50–70% of patients will heal or cope regardless of the nature and quality of the treatment rendered [73]. Thus, in the absence of accurate and precise measures of outcomes, a true understanding of the actual effect of treatment can easily be obscured.

19.5 Is Evidence-Based Medicine Really Evidence-Based in ACL Reconstruction?

A variety of Level I studies published on the topic of ACL reconstruction have focused on various patient-specific and surgical factors. Lubowitz et al. performed a meta-analysis comparing single- and double-bundle ACL reconstruction and did not find any difference [28, 43]. Foster et al. and Carey et al. compared allograft to autograft and found no difference [5, 11]. Holm et al. looked at hamstring versus patellar tendon autograft and found no difference [22]. Samuelsson looked at graft type and surgical technique and found no difference [60]. When looking at these studies, several causes for type II error can be identified, each of which are reviewed below.

19.6 Study Design

Basic science studies are extremely valuable in the field of orthopedic surgery. However, the pitfalls of improperly designed studies must be recognized. When in vitro testing parameters do not accurately reflect the in vivo condition, results may not necessarily be readily applicable to clinical scenarios. For example, most biomechanical studies on ACL reconstruction techniques only use testing systems that applied a small fraction of the load that the reconstructed knee would be exposed to in a living human subject performing high-level sports activities [37, 38, 46].

19.7 Patient Selection and Allocation of Treatment

The process of selecting patients and allocating them to a treatment groups is extremely important. Specifically, when comparing two groups of treatment, when possible, randomization should be applied to avoid baseline differences between the groups, which can confound the outcomes.

19.8 Selecting the Appropriate Outcome Measure

It is important to select the outcome tool that is most suitable and sensitive in detecting the primary outcome of a study as indicated in the hypothesis. Failure to do so will result in drawing an incorrect conclusion. An example is the use of KT arthrometer [MedMetric, San Diego, CA, USA] testing for knee instability after ACL reconstruction [70, 72]. Although this is an excellent tool for evaluation of anterior-posterior laxity, it has little value in assessing rotatory instability [70, 72]. Relying solely on the KT arthrometer to compare ACL reconstruction techniques is therefore not sufficient. Indeed it has been shown by more advanced biomechanical testing that nonanatomic techniques specifically fail to restore rotatory laxity, while they may be sufficient to restore anterior-posterior laxity [14].

It is important to select the outcome tool which is most suitable and sensitive in detecting the primary outcome of a study as indicated in the hypothesis. Failure to do so will result in drawing an incorrect conclusion.

A study on risk factors for the development of osteoarthritis after ACL reconstruction evaluated 50 patients at 6 years following surgery. The following ten potential factors involved in the development of osteoarthritis were assessed: meniscectomy, chondral damage, patellar tendon grafting, age at surgery, time delay between injury and surgery, type and intensity of post-surgery sport, quadriceps strength, hamstring strength, quadriceps-to-hamstring strength ratio, and residual joint laxity. However, despite the fact that the radiographs shown in the manuscript indicate nonanatomic tunnel placement [23], tunnel placement was not considered as a possible factor for osteoarthritis [30]. However, in accordance with

biomechanical studies, this should ideally have been evaluated (Fig. 19.1) [2].

A number of outcome tools—clinician-based and patient-reported—have been described to characterize the nature of patients' knee function [9, 19, 31, 51, 55, 61]. While many of these instruments have been validated in their ability to

detect knee-related disability, there is considerable variability in the types of patients that each of these measures were intended to assess (patients with knee osteoarthritis versus active, athletic patients with non-arthritic knee conditions) [39]. It should be noted that patient activity level is an important prognostic variable, which

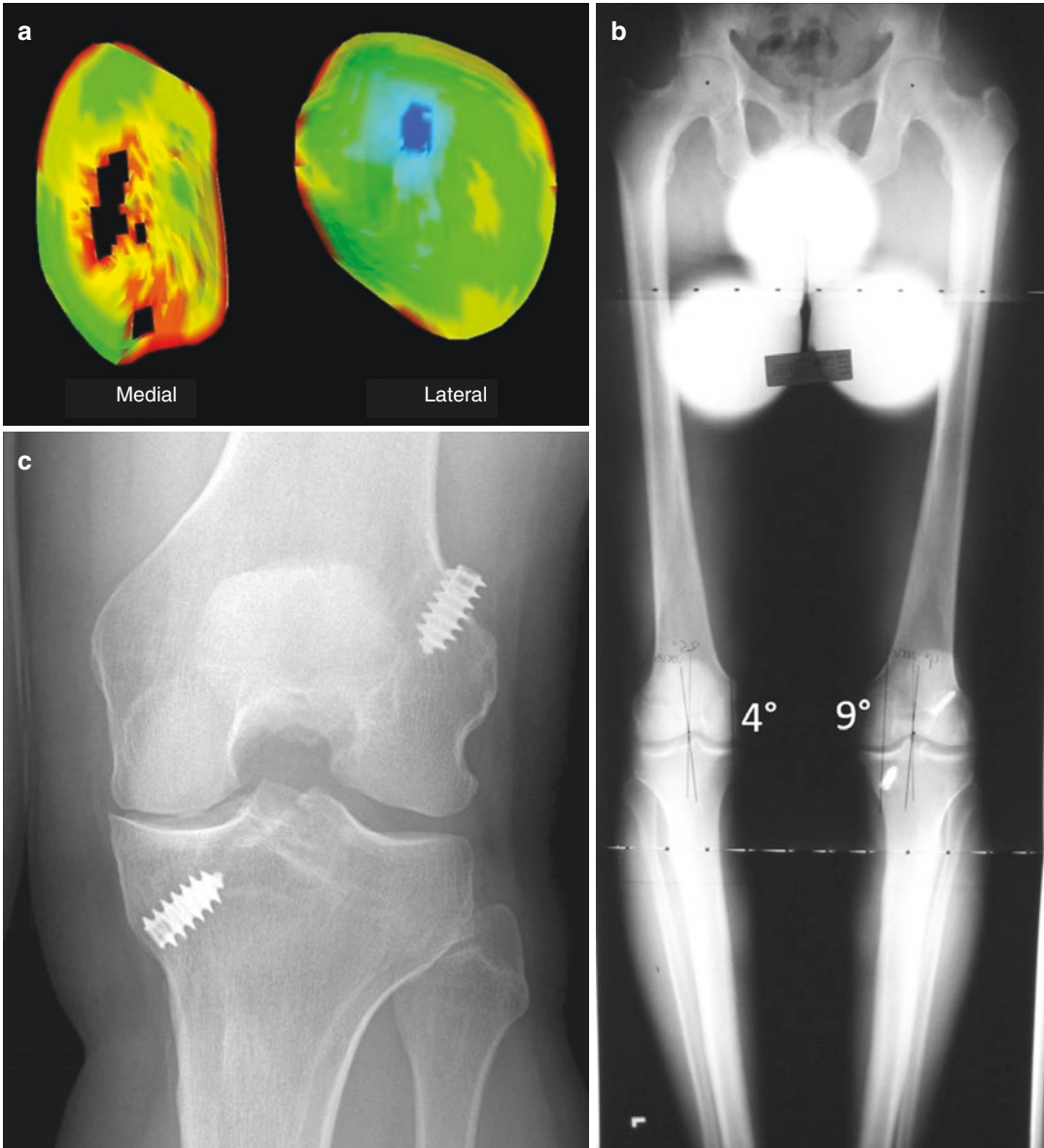


Fig. 19.1 (a) Pressure map of the articular cartilage of a left knee after nonanatomic ACL reconstruction showing increased pressures (red and black color) in the medial compartment. (b) Plain flexion posterior-anterior radio-

graph of a left knee after ACL reconstruction showing medial compartment osteoarthritis with joint space narrowing. (c) Long cassette radiograph of the same patient showing resulting varus malalignment [2]

does not always correlate to symptoms and function [39]. Consequently, studies should be designed in a way such that the outcome measures selected as dependent variables should be specific to the type of patients being studied. Thus far, this type of patient-specific outcome measure selection has been lacking in certain disciplines of orthopedic research [41].

The International Knee Documentation Committee (IKDC) has developed two rating scales, one “objective” and one “subjective” [20]. The first is clinician-based and grades patients as normal, nearly normal, abnormal, or severely abnormal on a variety of parameters that include effusion, motion, ligament laxity, crepitus, harvest-site pathology, radiographic findings, and single-leg hop test. The final patient grade is determined by the lowest grade in any given group. The subjective one is patient-reported and inquires about symptoms, sports activities, and ability to function, including stairs, squatting, running, and jumping. It has been demonstrated to be reliable, valid, and responsive when applied to a range of knee conditions, including ACL tears as well as meniscus and cartilage pathology [26, 27].

The Cincinnati Knee Rating System combines clinician-based evaluation with patient-reported symptoms and function [4, 63]. It is composed of 6 subscales that add up to 100 points: 20 for symptoms, 15 for daily and sports functional activities, 25 for physical examination, 20 for knee stability testing, 10 for radiographic findings, and 10 for functional testing [3]. It is most often employed to evaluate ACL injuries before and after reconstruction and is proven to be reliable, valid, and responsive [40, 55].

The modified Lysholm scale is a patient-related measure designed to evaluate outcomes after knee ligament surgery [36]. It is a questionnaire with eight items scaled to a maximum score of 100. Knee stability accounts for 25 points, pain for 25, locking for 15, and swelling and stair climbing for 10 each. In addition, a limp, use of a support, and squatting accounts for 5 points each [68]. It was originally developed in 1982 and later modified in 1985. This Lysholm score is one of the first outcome measures to rely on patient-

related symptoms and functions. It is frequently used in clinical research [21, 35].

The Single Assessment Numeric Evaluation (SANE) was designed specifically for college-age patients after ACL reconstruction [75]. It is very simple and involves just one question, asking a patient on a 0–100% scale how much of a percentage of normal they would rate their knee. Though application is easy, it is only useful when looking at a homogeneous cohort of patients who would interpret this one question similarly [39, 75].

The Knee Injury and Osteoarthritis Outcome Score (KOOS) is another patient-related measure. It consists of 5 separate scores: 9 questions for pain, 7 questions for symptoms, 17 questions for activities of daily living, 5 questions for sport and recreational function, and 4 items for quality of life [57]. The KOOS has been employed to evaluate ACL reconstruction but also meniscectomy, tibial osteotomy, and post-traumatic arthritis. The other benefit is that it has been validated in multiple languages [56, 58, 74, 78].

The quality-of-life outcome measure for chronic ACL deficiency was developed with input from ACL-deficient patients, primary care sports medicine physicians, orthopedic surgeons, athletic therapists, and physical therapists [39]. It consists of 31 visual analog questions relating to 5 categories: symptoms and physical complaints, work-related concerns, recreational activities and sports participation, lifestyle, and social and emotional health status relating to the knee [39, 45].

There are several instruments used to characterize patients' level of physical activity. The Tegner is probably the most popular in orthopedic publications. It aims to place a patient's activity level somewhere on a 0–10 scale based on their specific sport [68, 78]. However, although commonly used, this instrument has not been officially validated [41]. The Marx activity level is based on function-specific, rather than sport-specific, questions, and it incorporates the frequency of activities as well [41]. The scale consists of four questions, evaluating running, cutting, decelerating, and pivoting. Patients are asked to score frequency on a 0–4 scale for

each element for a total of 16 points. In contrast to the Tegner scale, the Marx scale has been validated [41].

Clinical Vignette

A 35-year-old female presents to the office complaining of right knee pain. She underwent right knee ACL reconstruction using a transtibial single-bundle technique 15 years ago. She has had no new injury. She states the knee feels stable. Examination in the office appears to confirm this with a 1A Lachman, negative anterior drawer, and guarding pivot shift test. She undergoes imaging with plain radiographs demonstrating a vertical tunnel orientation as well as an MRI indicating the ACL graft is intact but vertically oriented (Fig. 19.2). In addition, it reveals a degenerative medial meniscal tear. A CT scan was also ordered which reveals nonanatomic tunnel location (Fig. 19.2). She elects to undergo surgery for the meniscus tear. Examination under anesthesia reveals an exam very different from that in the office with a 2A Lachman, 1+ anterior drawer, and 2+ pivot shift test. Arthroscopy reveals a vertical graft, degenerative medial meniscal tear, and advanced degenerative changes of the medial compartment (Fig. 19.2). This clinical case illustrates that when insufficient or inappropriate outcome tools are used, clinical outcome is not accurately and reliably measured.

patient's compliance. The final grade attributed to amount of instability also results from a subjective judgment of the examiner. During the pre-course of ISAKOS, Osaka, in 2009, five selected experts were invited to perform a pivot shift test on a cadaveric lower body specimen. This setup was repeated at the Panther Global Summit in Pittsburgh in 2011 with 12 expert surgeons on an actual ACL-injured patient under general anesthesia (Fig. 19.3). During the exam, the pivot shift was quantified with an accelerometer. The results of this experiment showed no objective agreement [34]. This is partly due to differences among examiners, but also because the pivot shift is influenced not only by the ACL but also by the iliotibial band, capsule, medial meniscus, lateral meniscus, and bony morphology [48–50]. To make the pivot shift more objective, its use in conjunction with an accelerometer has been advocated and validated (Fig. 19.4) [1, 47].

Further objective outcome measures are currently being investigated and implemented in order to deal with variability in some of the more widely reported physical exam measures. Dynamic stereo in vivo radiography can be used for detailed kinematic analysis (Fig. 19.5) [65–67]. Three-dimensional computed tomography scanning is very valuable for evaluation of ACL tunnel placement [10, 32, 33]. Magnetic resonance imaging can be used for a variety of different purposes following ACL reconstruction (Fig. 19.6) such as evaluations of graft integrity, graft healing, inclination angle, tunnel position, as well as the status of the menisci and articular cartilage [7, 44].

19.9 Improving How We Measure Outcome

Although quantification is simple when using the aforementioned outcome and activity scales, the objective description of physical examination findings can be a challenging proposition. The pivot shift test, for example, is the most performed test to determine rotational instability in the knee, but it is dependent on the examiner skills and

19.10 Interpretation of Results

As described above, the clinical outcome following ACL reconstruction is frequently assessed using the IKDC scoring system. Although the use of this outcome tool has been validated, this was done using four specific categories: excellent, good, fair, and poor. However, various authors have inappropriately grouped the excellent and good outcome categories together, as well as the fair and poor category [5, 43]. This likely started

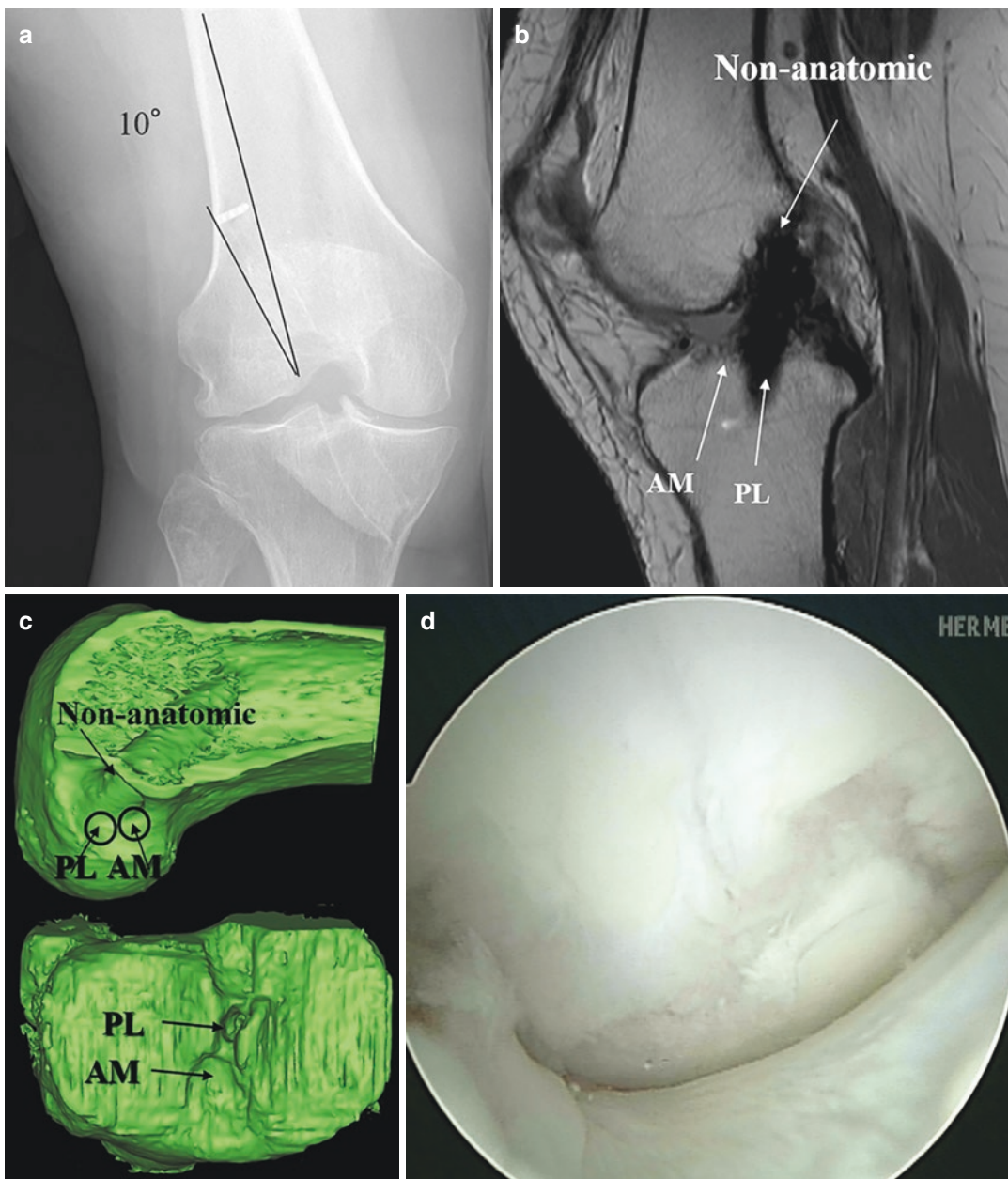


Fig. 19.2 A 35-year-old female, 15 years after right ACL reconstruction, presents with knee pain. Examination in the office appeared to suggest a stable knee. (a) Plain radiographs of a right knee showing status post-ACL reconstruction with a vertical tunnel orientation suggestive of nonanatomic tunnel placement. (b) MRI of the right knee

showing the ACL graft is intact but again vertical in orientation. (c) Three-dimensional CT scan of the same right knee confirming nonanatomic tunnel position. (d) Arthroscopy of that knee showing advanced degenerative changes in the medial compartment, frequently seen with residual rotatory instability after nonanatomic ACL reconstruction

as an attempt to dichotomize the results as normal/nearly normal versus abnormal/severely abnormal to allow for easier statistical analysis.

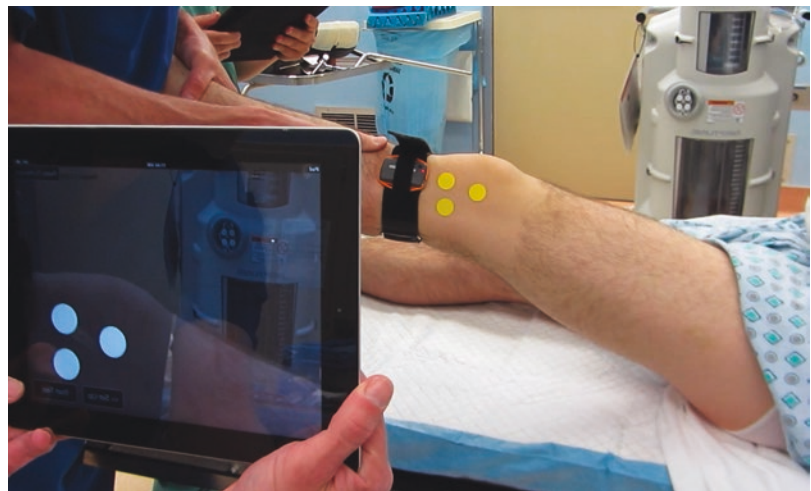
This fact may also be attributable to the concern that some of the components of the IKDC grading system, including the pivot shift, are subject-



Fig. 19.3 Panther Global Summit in Pittsburgh in 2011 with 12 expert surgeons performing the pivot shift test on a left knee of an actual ACL-injured patient under general anesthesia. During the exam, the pivot shift was quantified with an accelerometer, iPad app, and electro-

magnetic tracking system. The results of this experiment showed no objective agreement in the pivot shift grade among the experts and significantly improved agreement after instruction of a “standardized pivot shift test” [34]

Fig. 19.4 Examination of a left knee showing the application of an accelerometer and iPad application to quantify the pivot shift examination by the surgeon. Markers on the skin are used which are tracked with the iPad application [1, 47]



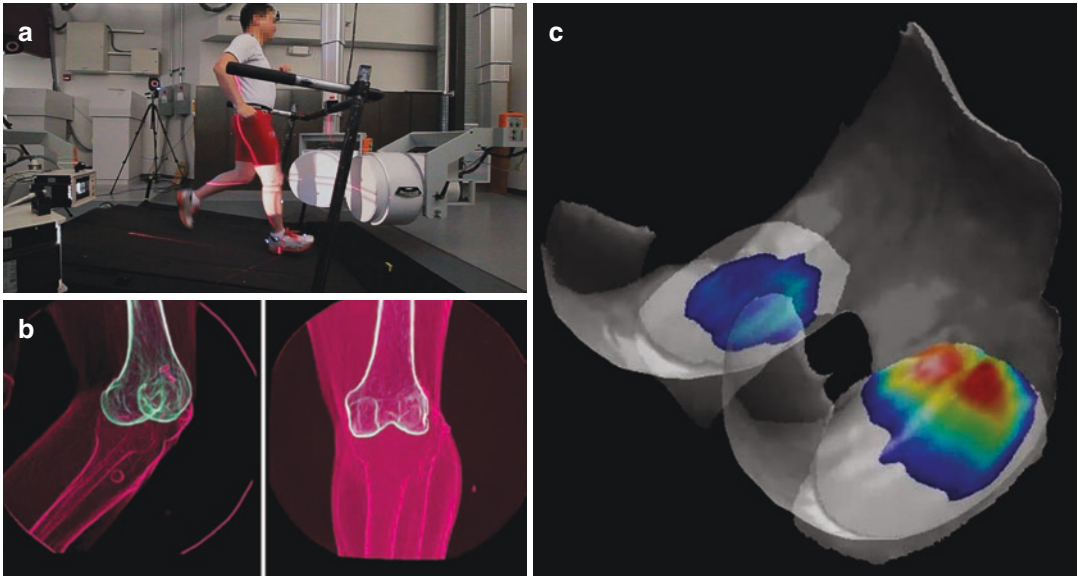


Fig. 19.5 In vivo dynamic stereo radiography of a left knee. (a) The patient runs on a treadmill, while real-time radiography is performed from orthogonal angles. (b) Videos are overlapped with a three-dimensional computed tomography scan of the knee such that it is possible to

track the distance and motion pattern between the tibia and femur as the patient runs and loads the knee. (c) Subsequent map of the knee outlining the distance and motion between the tibia and femur in real time [65–67]

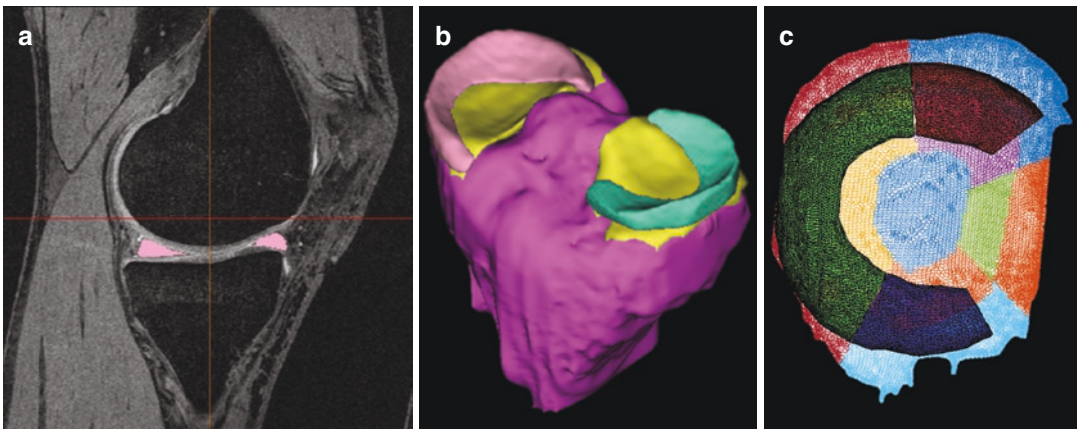


Fig. 19.6 High-resolution magnetic resonance imaging of a left knee. (a) Sagittal sequence outlining the medial meniscus. (b) Three-dimensional reconstruction of the magnetic resonance scan outlining the medial and lateral

meniscus. (c) Surface mapping of the anterior horn, body, and posterior horn of the medial meniscus and articular cartilage [7, 44]

tive in nature. Or perhaps some may believe that “nearly normal” is good enough when assessing clinical outcome after ACL reconstruction. However, although traditional methods to recon-

struct the ACL are not able to completely restore the knee to normal, this should be the goal of future reconstruction techniques [28, 43]. This point is illustrated by the case of a recent Level I

study that made headlines comparing the results of operative to nonoperative treatment of patients with ACL tears. Overall, there were 121 patients treated with either operative or nonoperative treatment [12]. Although randomization was performed, the crossover of patients was allowed per the intention to treat analysis. The authors concluded that no difference existed between the treatment groups, but when looking at the data in more detail, in the rehabilitation-only group, there were more meniscus injuries at final follow-up (13 vs. 1), more abnormal Lachman tests (75% vs. 35%), more abnormal pivot shift examinations (53% vs. 25%), and an increased total KT-arthrometer translation (8.3 vs. 6.6 mm). Thus, the authors' conclusions were based on a series of dependent variables that may not be the most relevant factors in ultimately determining success of treatment.

The use of reliable and valid outcome measure specific to the aim of the study is highly recommended when conducting a study. For example, for anatomic ACL reconstruction, a scoring system was developed, which was validated and tested for reliability and responsiveness in two separate studies [8, 69].

The use of reliable and valid outcome measure specific to the aim of the study is highly recommended when conducting a study.

19.11 Quality and Duration of Follow-Up

The quality of the follow-up is also extremely important. By the standard of publishing of major orthopedic journals with high impact factor, a follow-up percentage of at least 80% and ideally over 90% is warranted for a high-level prospective clinical trial. In addition, patients should be followed for at least 2 years following the tested procedure.

Two similar studies from these authors' institution highlight this difference in how outcomes are evaluated. The first study on the outcomes of

allograft ACL reconstruction was performed in 1996 and indicated a failure rate of 3% [18]. However, when outcomes of the same procedure were evaluated in 2011 with more advanced outcome measures, the failure rate was shown to approximate 15% [71].

Take-Home Message

- Finding no difference does not mean there is no difference.
- Orthopedists should be more like the cardiologists.
- Cardiology is leading in medicine with regard to research and innovation.
- In conducting research, it is important to form a sound hypothesis that allows for determination of the proper primary outcome measure for a particular study.
- The hypothesis of a research study dictates which statistical methods should be applied and avoids making statistical errors.
- It is important to select the outcome tool that is most suitable, reliable, valid, and sensitive in detecting the primary outcome of a study as indicated in hypothesis.
- Failure to do so will result in drawing an incorrect conclusion.
- Study design and the quality of the follow-up is also extremely important.
- Randomized studies with large numbers of patients and at least 80% follow-up at 2 years or longer are necessary to critically evaluate new surgical techniques.

References

1. Ahlden M, Araujo P, Hoshino Y, et al. Clinical grading of the pivot shift test correlates best with tibial acceleration. *Knee Surg Sports Traumatol Arthrosc.* 2012;20(4):708–12.
2. Andriacchi TP, Briant PL, Beville SL, Koo S. Rotational changes at the knee after ACL injury cause cartilage thinning. *Clin Orthop Relat Res.* 2006;442:39–44.
3. Barber-Westin SD, Noyes FR, McCloskey JW. Rigorous statistical reliability, validity, and responsiveness testing of the Cincinnati knee rating system in 350 subjects with uninjured, injured, or anterior cruciate ligament-reconstructed knees. *Am J Sports Med.* 1999;27(4):402–16.

4. Bollen S, Seedhom BB. A comparison of the Lysholm and Cincinnati knee scoring questionnaires. *Am J Sports Med.* 1991;19(2):189–90.
5. Carey JL, Dunn WR, Dahm DL, Zeger SL, Spindler KP. A systematic review of anterior cruciate ligament reconstruction with autograft compared with allograft. *J Bone Joint Surg Am.* 2009;91(9):2242–50.
6. Chavez I, Dorbecker N, Celis A. Direct intracardiac angiocardiology; its diagnostic value. *Am Heart J.* 1947;33(5):560–93.
7. Chu CR, Williams AA, West RV, et al. Quantitative magnetic resonance imaging UTE-T2* mapping of cartilage and meniscus healing after anatomic anterior cruciate ligament reconstruction. *Am J Sports Med.* 2014;42(8):1847–56. <https://doi.org/10.1177/0363546514532227>.
8. Desai N, Alentorn-Geli E, van Eck CF, et al. A systematic review of single-versus double-bundle ACL reconstruction using the anatomic anterior cruciate ligament reconstruction scoring checklist. *Knee Surg Sports Traumatol Arthrosc.* 2016;24(3):862–72.
9. Eastlack ME, Axe MJ, Snyder-Mackler L. Laxity, instability, and functional outcome after ACL injury: copers versus noncopers. *Med Sci Sports Exerc.* 1999;31(2):210–5.
10. Forsythe B, Kopf S, Wong AK, et al. The location of femoral and tibial tunnels in anatomic double-bundle anterior cruciate ligament reconstruction analyzed by three-dimensional computed tomography models. *J Bone Joint Surg Am.* 2010;92(6):1418–26.
11. Foster TE, Wolfe BL, Ryan S, Silvestri L, Kaye EK. Does the graft source really matter in the outcome of patients undergoing anterior cruciate ligament reconstruction? An evaluation of autograft versus allograft reconstruction results: a systematic review. *Am J Sports Med.* 2010;38(1):189–99.
12. Frobell RB, Roos EM, Roos HP, Ranstam J, Lohmander LS. A randomized trial of treatment for acute anterior cruciate ligament tears. *N Engl J Med.* 2010;363(4):331–42.
13. Fujimoto E, Sumen Y, Deie M, Yasumoto M, Kobayashi K, Ochi M. Anterior cruciate ligament graft impingement against the posterior cruciate ligament: diagnosis using MRI plus three-dimensional reconstruction software. *Magn Reson Imaging.* 2004;22(8):1125–9.
14. Gabriel MT, Wong EK, Woo SL, Yagi M, Debski RE. Distribution of in situ forces in the anterior cruciate ligament in response to rotatory loads. *J Orthop Res.* 2004;22(1):85–9.
15. Gillquist J, Messner K. Anterior cruciate ligament reconstruction and the long-term incidence of gonarthrosis. *Sports Med.* 1999;27(3):143–56.
16. Harner CD, Irrgang JJ, Paul J, Dearwater S, Fu FH. Loss of motion after anterior cruciate ligament reconstruction. *Am J Sports Med.* 1992;20(5):499–506.
17. Harner CD, Marks PH, Fu FH, Irrgang JJ, Silby MB, Mengato R. Anterior cruciate ligament reconstruction: endoscopic versus two-incision technique. *Arthroscopy.* 1994;10(5):502–12.
18. Harner CD, Olson E, Irrgang JJ, Silverstein S, Fu FH, Silbey M. Allograft versus autograft anterior cruciate ligament reconstruction: 3- to 5-year outcome. *Clin Orthop Relat Res.* 1996;324:134–44.
19. Heckman JD. Are validated questionnaires valid? *J Bone Joint Surg Am.* 2006;88(2):446.
20. Hefti F, Muller W. [Current state of evaluation of knee ligament lesions. The new IKDC knee evaluation form]. *Orthopade.* 1993;22(6):351–62.
21. Hoher J, Munster A, Klein J, Eypasch E, Tiling T. Validation and application of a subjective knee questionnaire. *Knee Surg Sports Traumatol Arthrosc.* 1995;3(1):26–33.
22. Holm I, Oiestad BE, Risberg MA, Aune AK. No difference in knee function or prevalence of osteoarthritis after reconstruction of the anterior cruciate ligament with 4-strand hamstring autograft versus patellar tendon-bone autograft: a randomized study with 10-year follow-up. *Am J Sports Med.* 2010;38(3):448–54.
23. Illingworth KD, Hensler D, Working ZM, Macalena JA, Tashman S, Fu FH. A simple evaluation of anterior cruciate ligament femoral tunnel position: the inclination angle and femoral tunnel angle. *Am J Sports Med.* 2011;39(12):2611–8.
24. Iriuchishima T, Tajima G, Ingham SJ, et al. Intercondylar roof impingement pressure after anterior cruciate ligament reconstruction in a porcine model. *Knee Surg Sports Traumatol Arthrosc.* 2009;17(6):590–4.
25. Iriuchishima T, Tajima G, Ingham SJ, Shen W, Smolinski P, Fu FH. Impingement pressure in the anatomical and nonanatomical anterior cruciate ligament reconstruction: a cadaver study. *Am J Sports Med.* 2010;38(8):1611–7.
26. Irrgang JJ, Anderson AF. Development and validation of health-related quality of life measures for the knee. *Clin Orthop Relat Res.* 2002;402:95–109.
27. Irrgang JJ, Anderson AF, Boland AL, et al. Responsiveness of the International Knee Documentation Committee Subjective Knee Form. *Am J Sports Med.* 2006;34(10):1567–73.
28. Irrgang JJ, Bost JE, Fu FH. Re: outcome of single-bundle versus double-bundle reconstruction of the anterior cruciate ligament: a meta-analysis. *Am J Sports Med.* 2009;37(2):421–2; author reply 422.
29. Johnson DL, Swenson TM, Irrgang JJ, Fu FH, Harner CD. Revision anterior cruciate ligament surgery: experience from Pittsburgh. *Clin Orthop Relat Res.* 1996;325:100–9.
30. Keays SL, Newcombe PA, Bullock-Saxton JE, Bullock MI, Keays AC. Factors involved in the development of osteoarthritis after anterior cruciate ligament surgery. *Am J Sports Med.* 2010;38(3):455–63.
31. Kocher MS, Steadman JR, Briggs K, Zurakowski D, Sterett WI, Hawkins RJ. Determinants of patient satisfaction with outcome after anterior cruciate ligament reconstruction. *J Bone Joint Surg Am.* 2002;84-A(9):1560–72.

32. Kopf S, Forsythe B, Wong AK, et al. Nonanatomic tunnel position in traditional transtibial single-bundle anterior cruciate ligament reconstruction evaluated by three-dimensional computed tomography. *J Bone Joint Surg Am.* 2010;92(6):1427–31.
33. Kopf S, Forsythe B, Wong AK, Tashman S, Irrgang JJ, Fu FH. Transtibial ACL reconstruction technique fails to position drill tunnels anatomically in vivo 3D CT study. *Knee Surg Sports Traumatol Arthrosc.* 2012;20(11):2200–7.
34. Kopf S, Musahl V, Perka C, Kauert R, Hoburg A, Becker R. The influence of applied internal and external rotation on the pivot shift phenomenon. *Knee Surg Sports Traumatol Arthrosc.* 2017;25(4):1106–10.
35. Lukianov AV, Gillquist J, Grana WA, DeHaven KE. An anterior cruciate ligament (ACL) evaluation format for assessment of artificial or autologous anterior cruciate reconstruction results. *Clin Orthop Relat Res.* 1987;218:167–80.
36. Lysholm J, Gillquist J. Evaluation of knee ligament surgery results with special emphasis on use of a scoring scale. *Am J Sports Med.* 1982;10(3):150–4.
37. Markolf KL, Feeley BT, Jackson SR, McAllister DR. Biomechanical studies of double-bundle posterior cruciate ligament reconstructions. *J Bone Joint Surg Am.* 2006;88(8):1788–94.
38. Markolf KL, Park S, Jackson SR, McAllister DR. Simulated pivot-shift testing with single and double-bundle anterior cruciate ligament reconstructions. *J Bone Joint Surg Am.* 2008;90(8):1681–9.
39. Marx RG. Knee rating scales. *Arthroscopy.* 2003;19(10):1103–8.
40. Marx RG, Jones EC, Allen AA, et al. Reliability, validity, and responsiveness of four knee outcome scales for athletic patients. *J Bone Joint Surg Am.* 2001;83-A(10):1459–69.
41. Marx RG, Stump TJ, Jones EC, Wickiewicz TL, Warren RF. Development and evaluation of an activity rating scale for disorders of the knee. *Am J Sports Med.* 2001;29(2):213–8.
42. Mayo Robson AW. Ruptured crucial ligaments and their repair by operation. *Ann Surg.* 1903;37(5):716–8.
43. Meredith RB, Vance KJ, Appleby D, Lubowitz JH. Outcome of single-bundle versus double-bundle reconstruction of the anterior cruciate ligament: a meta-analysis. *Am J Sports Med.* 2008;36(7):1414–21.
44. Miyawaki M, Hensler D, Illingworth KD, Irrgang JJ, Fu FH. Signal intensity on magnetic resonance imaging after allograft double-bundle anterior cruciate ligament reconstruction. *Knee Surg Sports Traumatol Arthrosc.* 2014;22(5):1002–8.
45. Mohtadi N. Development and validation of the quality of life outcome measure (questionnaire) for chronic anterior cruciate ligament deficiency. *Am J Sports Med.* 1998;26(3):350–9.
46. Morimoto Y, Ferretti M, Ekdahl M, Smolinski P, Fu FH. Tibiofemoral joint contact area and pressure after single- and double-bundle anterior cruciate ligament reconstruction. *Arthroscopy.* 2009;25(1):62–9.
47. Musahl V, Griffith C, Irrgang JJ, et al. Validation of quantitative measures of rotatory knee laxity. *Am J Sports Med.* 2016;44(9):2393–8.
48. Musahl V, Hoshino Y, Ahlden M, et al. The pivot shift: a global user guide. *Knee Surg Sports Traumatol Arthrosc.* 2012;20(4):724–31.
49. Musahl V, Hoshino Y, Becker R, Karlsson J. Rotatory knee laxity and the pivot shift. *Knee Surg Sports Traumatol Arthrosc.* 2012;20(4):601–2.
50. Musahl V, Rahnama-Azar AA, Costello J, et al. The influence of meniscal and anterolateral capsular injury on knee laxity in patients with anterior cruciate ligament injuries. *Am J Sports Med.* 2016;44(12):3126–31.
51. Neeb TB, Aufdemkampe G, Wagener JH, Mastenbroek L. Assessing anterior cruciate ligament injuries: the association and differential value of questionnaires, clinical tests, and functional tests. *J Orthop Sports Phys Ther.* 1997;26(6):324–31.
52. Nishimori M, Sumen Y, Sakaridani K, Nakamura M. An evaluation of reconstructed ACL impingement on PCL using MRI. *Magn Reson Imaging.* 2007;25(5):722–6.
53. Passler HH. The history of the cruciate ligaments: some forgotten (or unknown) facts from Europe. *Knee Surg Sports Traumatol Arthrosc.* 1993;1(1):13–6.
54. Powell WJ Jr, Wittenberg J, Dinsmore RE, Miller SW, Maturi RA. Definition of cardiac structures using computerized tomography in isolated arrested and beating canine hearts. *Am J Cardiol.* 1977;39(5):690–6.
55. Risberg MA, Holm I, Steen H, Beynnon BD. Sensitivity to changes over time for the IKDC form, the Lysholm score, and the Cincinnati knee score. A prospective study of 120 ACL reconstructed patients with a 2-year follow-up. *Knee Surg Sports Traumatol Arthrosc.* 1999;7(3):152–9.
56. Roos EM, Ostenberg A, Roos H, Ekdahl C, Lohmander LS. Long-term outcome of meniscectomy: symptoms, function, and performance tests in patients with or without radiographic osteoarthritis compared to matched controls. *Osteoarthr Cartil.* 2001;9(4):316–24.
57. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. *J Orthop Sports Phys Ther.* 1998;28(2):88–96.
58. Roos EM, Roos HP, Ryd L, Lohmander LS. Substantial disability 3 months after arthroscopic partial meniscectomy: a prospective study of patient-relevant outcomes. *Arthroscopy.* 2000;16(6):619–26.
59. Rosenberg T. Techniques for ACL reconstruction with Multi-Trac drill guide. Mansfield: Accuflex Microsurgical Inc.; 1994.
60. Samuelsson K, Andersson D, Karlsson J. Treatment of anterior cruciate ligament injuries with special reference to graft type and surgical technique: an assess-

- ment of randomized controlled trials. *Arthroscopy*. 2009;25(10):1139–74.
61. Sernert N, Kartus J, Kohler K, et al. Analysis of subjective, objective and functional examination tests after anterior cruciate ligament reconstruction. A follow-up of 527 patients. *Knee Surg Sports Traumatol Arthrosc*. 1999;7(3):160–5.
 62. Sgaglione NA. Revision ACL reconstruction. Presented at Orthopedics Today Hawaii 2011. Jan 16–19 Koloa, Hawaii; 2011.
 63. Sgaglione NA, Del Pizzo W, Fox JM, Friedman MJ. Critical analysis of knee ligament rating systems. *Am J Sports Med*. 1995;23(6):660–7.
 64. Snook GA. A short history of the anterior cruciate ligament and the treatment of tears. *Clin Orthop Relat Res*. 1983;172:11–3.
 65. Tashman S, Anderst W. In-vivo measurement of dynamic joint motion using high speed biplane radiography and CT: application to canine ACL deficiency. *J Biomech Eng*. 2003;125(2):238–45.
 66. Tashman S, Collon D, Anderson K, Kolowich P, Anderst W. Abnormal rotational knee motion during running after anterior cruciate ligament reconstruction. *Am J Sports Med*. 2004;32(4):975–83.
 67. Tashman S, Kolowich P, Collon D, Anderson K, Anderst W. Dynamic function of the ACL-reconstructed knee during running. *Clin Orthop Relat Res*. 2007;454:66–73.
 68. Tegner Y, Lysholm J. Rating systems in the evaluation of knee ligament injuries. *Clin Orthop Relat Res*. 1985;198:43–9.
 69. van Eck CF, Gravare-Silbernagel K, Samuelsson K, et al. Evidence to support the interpretation and use of the anatomic anterior cruciate ligament reconstruction checklist. *J Bone Joint Surg Am*. 2013;95(20):e1531–9.
 70. van Eck CF, Loopik M, van den Bekerom MP, Fu FH, Kerkhoffs GM. Methods to diagnose acute anterior cruciate ligament rupture: a meta-analysis of instrumented knee laxity tests. *Knee Surg Sports Traumatol Arthrosc*. 2013;21(9):1989–97.
 71. van Eck CF, Schkrohowsky JG, Working ZM, Irrgang JJ, Fu FH. Prospective analysis of failure rate and predictors of failure after anatomic anterior cruciate ligament reconstruction with allograft. *Am J Sports Med*. 2012;40(4):800–7.
 72. van Eck CF, van den Bekerom MP, Fu FH, Poolman RW, Kerkhoffs GM. Methods to diagnose acute anterior cruciate ligament rupture: a meta-analysis of physical examinations with and without anaesthesia. *Knee Surg Sports Traumatol Arthrosc*. 2013;21(8):1895–903.
 73. Vavken P. Rationale for and methods of superiority, noninferiority, or equivalence designs in orthopaedic, controlled trials. *Clin Orthop Relat Res*. 2011;469(9):2645–53.
 74. W-Dahl A, Toksvig-Larsen S, Roos EM. A 2-year prospective study of patient-relevant outcomes in patients operated on for knee osteoarthritis with tibial osteotomy. *BMC Musculoskelet Disord*. 2005;6:18.
 75. Williams GN, Taylor DC, Gangel TJ, Uhorchak JM, Arciero RA. Comparison of the single assessment numeric evaluation method and the Lysholm score. *Clin Orthop Relat Res*. 2000;373:184–92.
 76. Wilson FN, Johnston FD, Hill IGW, Macleod AG, Barker PS. The significance of electrocardiograms characterized by an abnormally long QRS interval and by broad S deflections in Lead I. *Am Heart J*. 1934;9:459.
 77. Wood EH, Ritman EL, Robb RA, Harris LD, Ruegsegger P. Noninvasive numerical vivisection of anatomic structure and function of the intact circulatory system using high temporal resolution cylindrical scanning computerized tomography. *Med Instrum*. 1977;11(3):153–9.
 78. Wright RW. Knee injury outcomes measures. *J Am Acad Orthop Surg*. 2009;17(1):31–9.

Stephen Lyman

20.1 Power

Power is a powerful word. It's extremely flexible with no fewer than nine definitions according to the Merriam-Webster Dictionary [6]. The final definition is the one we're referring to in this chapter: "the probability of rejecting the null hypothesis in a statistical test when a particular alternative hypothesis happens to be true." In clinical research, power refers to the ability to detect a difference in diagnostic, prognostic, or treatment effectiveness if one exists. This is a primarily theoretical construct but has practical implications. It applies to any study design in which you test a hypothesis whether it be between two different groups of research subjects given a diagnostic test, prognostic evaluation, or treatment regimen or within the same subjects before and after an intervention [4].

20.2 Does Power Matter?

Statistical power provides investigators and readers with a sense of whether the study actually answers the research question of interest. Adequately powered studies can help improve our ability to diagnose and manage orthopedic

Clinical Vignette 1

Consider a research study comparing two alternative surgical approaches for patients suffering from a rare orthopedic condition with just a handful of subjects available for study (Table 20.1, Example 1). Assuming we have equipoise in not knowing which treatment is more effective and/or safe, we would be justified in randomizing patients presenting with this condition to either of the two treatment groups. In order to maximize power (and efficiency), we randomize them in a 1:1 allocation. Preoperatively we assess their pain levels on a 100 mm visual analogue scale. We randomize patients to receive Surgery A or Surgery B. This randomization works, and we find that the pre-treatment pain levels are equivalent between the two groups of patients ($p = 0.87$). Six weeks after treatment, we measure the patients' pain levels again. At this time we find that the patients who underwent Surgery A have a pain score of 34, while patients who underwent Surgery B have a pain score of 52 (22 point difference between groups). Surgery A seems to be more effective in reducing pain in these patients, right? Not so fast. First we must perform a statistical test to determine if the difference between treatment groups is statistically significant. To our surprise the test result's p-value comes back a nonsignificant 0.29.

S. Lyman (✉)
Hospital for Special Surgery, New York, NY, USA
e-mail: lymans@hss.edu

Clinical Vignette 2

Now imagine another study in which we compare two different surgical interventions for a common orthopedic condition where we're able to recruit a large number of patients (Table 20.1, Example 2). Again we randomize the patients to treatment group—this time Surgery C versus Surgery D. The randomization again works, and we find that patients receiving each treatment had similar pre-treatment pain levels ($p = 0.37$). Six weeks after surgery, we find that both groups have substantially less pain than before surgery ($p < 0.01$). The group receiving Surgery C has less pain, but the difference is only 2 mm out of a 100 mm pain score. This time the statistical test results in a p -value of <0.0001 , which is highly statistically significant using the usual critical p -value criteria of <0.05 . Yet there is only a slight difference between the group means, suggesting little clinical significance.

conditions. But studies can be underpowered or overpowered, as the examples below show:

The findings for both of these examples are completely the result of statistical power. A study with a very small sample size may show a very large difference in results between two interventions, but a statistical test of that difference may not detect significance. Conversely, a study with a very large sample size may find a statistically significant difference between the effects of two interventions, but the difference may be clinically irrelevant. The first example is an underpowered study. The second is overpowered.

Underpowering a study may result in missing a true effect simply because not enough subjects were enrolled in the study. Underpowering usually results in a null finding even despite the presence of a true effect. We miss an opportunity to improve our understanding of clinical care, and our patients are worse off as a result. If a study is underpowered, researchers must be cautious in interpreting their results, while readers must con-

Table 20.1 Examples of underpowered and overpowered studies

	Example #1		Example #2	
	Surgery A	Surgery B	Surgery C	Surgery D
Sample size	5	5	2000	2000
Mean presurgical pain \pm SD	88 \pm 17	86 \pm 19	85 \pm 38	86 \pm 33
Mean postsurgical pain \pm SD	34 \pm 25	52 \pm 26	38 \pm 13	40 \pm 14
Between group p -value	0.29		<0.0001	
Statistical power	20%		$>99\%$	

sider whether the findings reflect the truth or are the result of inadequate power. Because of these concerns, editors may decline publishing underpowered studies.

Overpowered studies may yield findings that are statistically significant but clinically irrelevant. Using outcome measures with known minimal clinically important change (MCIC) or other well-established benchmarks of clinical effectiveness can be used to avoid misinterpreting overpowered studies.

Overpowering a study may be a waste of resources but may also allow for subgroup analyses. In Example 2, a subgroup analysis focusing on a specific age range of patients may reveal a bigger, clinically significant difference between treatments in younger patients, despite a small difference in the overall study population. A study overpowered to test the main hypothesis may have enough power to detect meaningful differences in subgroups of patients.

Fact Box 20.1

Statistical power determines your ability to detect a difference if one truly exists. Overpowered studies may yield findings that are statistically significant but clinically irrelevant. Underpowering a study may result in missing a true effect simply because not enough subjects were enrolled in the study.

20.3 Why Is Power Needed?

Statistical power is necessary anytime you want to test hypotheses: to determine if there is a statistically significant difference between groups or a statistically significant relationship between two variables [4]. *Statistical power determines your ability to detect a difference if one truly exists.*

This rationale is both scientific and philosophical. When conducting research, the project is only worth performing if it is possible to reject the null hypothesis. Underpowered research is less likely to be published or to contribute meaningfully to improving our understanding of the physical world. Therefore, *it is unethical (philosophical argument) to perform underpowered studies, because humans are being subject to unnecessary experimentation and risk of harm.* Furthermore, a null finding from an underpowered study may incorrectly be interpreted as evidence that the medical intervention studied has no benefit.

Fact Box 20.2

It is unethical to perform underpowered studies, because humans are being subject to unnecessary experimentation and risk of harm.

If you are not formally testing hypotheses, then statistical power is not strictly necessary. However, if you are interested in assessing the correlation between two variables, then you still need adequate sample size. For example, if you were evaluating whether ultrasound could diagnose a collateral ligament tear as well as a more expensive imaging modality, then you need a sample size large enough to calculate a reliable estimate of the ability of ultrasound to correctly diagnose the tear. A study of three subjects evaluated with both MRI and ultrasound would be inadequate to answer this question. There are only 4 possible findings: 0% accuracy (0 of 3 ultrasounds agree with MRI), 33% accuracy (1 of

3 agree), 67% accuracy, and 100% accuracy. Obviously, a much larger sample size would be necessary to establish meaningful estimate of ultrasound diagnostic utility in this setting. An entire literature has been developed evaluating the sample size requirements of reliability studies [3, 7, 8, 10].

20.4 Properties of Power

Statistical power is customarily represented as a percentage between 0 and 100% or sometimes as a proportion between 0 and 1. *It tells you what your chances are of missing a true finding.* Power is calculated as $1 - \beta$ (beta) where β is the Type II error or the likelihood of rejecting the alternative hypothesis if true. If β were 0.2, then $1 - \beta$ would be 0.80, or 80% power, to detect a difference if it truly exists. For most clinical studies, 80% power is considered the lowest acceptable power value since this means you have just 1 in 5 chance of missing a true finding. When considering interventions or testing hypotheses for which there are serious consequences, a power of 90% (1 in 10 chance of missing a true finding) or even higher may be warranted.

Let's imagine an established surgical procedure that is highly effective and relatively inexpensive but has high risk of adverse events (e.g., perioperative fracture). A new alternative surgical technique is believed to be both effective and safer than the established procedure but costs substantially more. In comparing these two surgical interventions, we would not want to miss a

Fact Box 20.3

Power tells you what your chances are of missing a true finding and is determined by sample size, variability, frequency, the critical p -value, and the minimum relevant effect size. The best approach is to set your p -value and effect size and estimate your variability and/or frequency and then calculate the sample size you need for 80% (or 90% power).

true treatment effect if it exists, so we might consider powering our study to 90% or higher. If we missed a true treatment effect in the new technique's favor, it might not be adopted into clinical practice due to the high costs despite being safer and more effective.

Power is determined by sample size, variability, frequency, the critical p -value, and the minimum relevant effect size. Adjusting any of these characteristics changes the statistical power.

Sample size is what most scientists think of first when considering statistical power [1]. The higher the sample size, the higher the power all else remaining equal. Conversely, a smaller sample size will always have a less power all else being equal. This is also the most easily modifiable factor in calculated expected power. We can usually recruit more patients, but other components of power are often more difficult or impossible to change.

Variability is a measure of how much spread exists in the variables being considered. A variable that is highly variable (pun intended) between study subjects will result in a larger standard deviation. The larger the variation, the more subjects will be needed, because a larger variation will make any difference between the groups harder to detect. Variability only applies to power calculations for continuous or scale parameters. Frequency (see below) is used for discrete variables.

A study's power is optimized when the **frequency** of either a discrete outcome or explanatory variable is balanced. A study being powered on a binary (or ordinal) variable with very low frequency will require many more subjects in order to achieve adequate statistical power than a study in which the frequency is balanced across groups. For example, if 50% of the study subjects are female and 50% are male, then an analysis comparing sex differences would have optimum power. However, if intersex (those with biological reproductive anatomy that are not fully male or fully female) were of interest as a third category of sex (an estimated 1% of live births) [2], then this low frequency group would have pro-

found implications for statistical power for a study considering sex differences.

The **p -value** represents your chance of a detected effect not being true. A critical **p -value** of 0.05 is usually considered acceptable for clinical research. This represents less than a 1 in 20 chance of falsely rejecting the null hypothesis if the null hypothesis is true. A smaller p -value may be desired if 1 in 20 seems too large an uncertainty. In that case a critical p -value of 0.01 (1 in 100) or even 0.001 (1 in 1000) may be warranted. These more certain critical p -values will decrease power all else being equal. Conversely, if a larger p -value were considered acceptable (e.g., p of 0.1 or 1 in 10), power would be increased all else being equal. p -Values are not usually modified unless there is a strong justification to be more or less refined in what is considered significant.

When many different comparisons are being conducted within the same study, p -values will often be adjusted for multiple comparisons. One of the most well-known adjustments is the Bonferroni correction in which the critical p -value of 0.05 is divided by the number of comparisons being made. If there were 10 hypotheses being tested, the new critical p -value would become 0.005 ($0.05 \div 10$ comparisons). The power would then be calculated based on this new effective p -value. As you can imagine, the Bonferroni correction can be extremely conservative when a large number of comparisons are planned.

Effect size refers to the magnitude of effect (difference between groups) you expect to find. Best practice is to use the smallest effect size deemed clinically relevant. If you do not have an expected effect size based on previous information (e.g., pilot data or previously published studies), then using your clinical judgment may be required but could be difficult to justify in this era of data-driven information.

Special consideration should be given to effect sizes related to patient-reported outcome measures (PROMs), which are very common outcome tools in elective orthopedic research. We often use the minimal clinically important difference

(MCID), which is also sometimes called the minimal clinically important change (MCIC) or minimal clinically important improvement (MCII). All of these are slightly different concepts, but we will use them interchangeably here as the minimum effect size we'd like to be able to detect when comparing two groups of patients using their PROM scores [5]. Conceptually, the MCIC is the smallest change in PROM score for which a subject can actually discern a difference in their state of health. This concept is particularly useful when you do not have previous data on which to estimate your effect size. A distribution-based MCIC is usually considered $0.5 \times$ standard deviation of the PROM score, which gives you a rough estimate of the MCIC. Some recent work has suggested that this distribution-based MCIC calculation is actually closer to the concept of minimal detectable change (MDC), which is essentially the calibration variability of the PROM [9]. However, when previous anchor-based MCICs are not available, the distribution-based method is usually the only alternative.

As a rule of thumb, the MCIC would be the smallest change expected to make a difference. This difference may be in a subject's health, quality of life, satisfaction, or a myriad of other measures considered clinically important. When using the distribution-based approach, this is often much smaller than we may expect from a treatment thought to be effective. If true, this will result in an overpowering of the study but may also allow for subgroup analyses to determine in which patients the treatment is most effective (or ineffective), a concept often called heterogeneity of treatment effect.

Adjusting any of these factors will change the power for a given study. Since most peer-reviewed journals require a p -value of 0.05 or less to be considered statistically significant, this is the power characteristic least easily modifiable for a power calculation. Although powering to 0.01 or 0.001 is not unheard of, powering to a p -value of larger than 0.05 is usually not acceptable.

Variability and frequency can be adjusted through study design considerations. Variability can

be reduced by having narrow inclusion criteria, which creates a more homogenous study population. The trade-off here is that you may be limiting the generalizability of the results of the study. For example, if you isolate your study to teenage female soccer players, you cannot extend your results very easily to college-age male basketball players or other athletic populations, even if you've managed to reduce variability in your study population.

Investigators may be tempted to play fast and loose with effect sizes to tweak a power calculation, but this should be done with caution. If you use too large an effect size, you'll have what appears to be an adequately powered study, but you may be hopelessly underpowered to detect the effect size you are likely to discover even if the treatment is reasonably effective. You'd still end up with a negative result even though the effect size revealed was clinically meaningful.

Therefore, adjusting sample size is usually the most practical approach to achieving sufficient statistical power, and this is why we often equate sample size with power. *The best approach is to set your p -value and effect size, and estimate your variability and/or frequency and then calculate the sample size you need for 80% (or 90% power).* Another approach is to set all those parameters, choose a practical target sample size, and see if your power reaches 80% (or higher). If not, you can tweak your sample size upward until you reach your target power.

Not uncommonly, you have very few cases (e.g., infected total knee arthroplasty (TKA) cases) but a nearly unlimited number of potential control subjects (noninfected TKA cases). In this case, you can identify all of your infected TKA cases and then match multiple control subjects per case. Statistical power will be increased for each additional control per case you add. Many case-control studies are performed with a 1:1 control/case ratio to maximize efficiency, but when power is limited due to the small number of cases available, efficiency can be sacrificed by matching 2:1, 3:1, or even 4:1. Practically speaking, power gains are minimal after 6:1, so it's really not worth going higher than that.

A final consideration in all of this is your analytic plan. The statistical tests proposed also determine your likely power, though this goes beyond the scope of this chapter and should be discussed with a statistician.

20.5 Is Statistical Power Ever Truly Known?

Statistical power as a “truth” is unknowable. Rather, it is an estimate of the likelihood of finding a true difference if one exists, and it is only as accurate as the inputs we enter into our power calculation. If our estimates of effect size are too low, we may be overpowered for the actual effect size found. That’s not problematic except that the study was less efficient than it could have been. A worse scenario would be if you were overly optimistic with your effect size and end with a null study. Likewise, if your variability is higher than expected, or the frequency of one of your groups is less common than expected, you’ll lose power. Only your p -value and sample size are stable targets in a power calculation. Moreover, difficulty with patient recruitment or loss to follow up in prospective studies can negatively affect sample size. To account for these possibilities, it is customary to calculate your power and sample size and then artificially inflate the sample size by the expected dropout rate (often a 10–20% inflation factor if not higher).

A well-managed institutional review board (IRB, ethics review panel) will require an a priori (before starting the study) power calculation for all clinical research. Many peer-reviewed medical journals also require these a priori calculations. In cases where a study team failed to calculate power a priori, a post hoc power calculation is often appropriate to assure themselves, journal reviewers, and eventual readers that the study was adequately powered. Even in the case of research where an a priori power calculation was performed, if the estimates used in the original calculation were inaccurate, a post hoc power calculation can be reassuring to the investigators, reviewers, and readers.

20.6 How to Calculate Statistical Power?

Until a few decades ago, power calculations were very time-consuming, hand-performed calculations that would keep statisticians busy until late into the night. Today most statistical software programs provide power calculation tools. Both stand-alone power calculation programs and macros are available for common statistical packages such as SAS or R.

Early free web-based power calculators were proven unreliable. Though some have been improved over time, they are still use-at-your-own-risk since the underlying calculation is a bit of a black box, and it’s impossible to know if it was coded properly. Professional statistical software packages have much more rigorous design and testing procedures in place to give you more confidence in your power calculation.

Ultimately, for the clinician undertaking clinical research, the safest way to be assured that your power calculation is performed correctly is by consulting with a statistician. Statisticians rely on you for your clinical expertise. You should rely on them for their statistical expertise. However, if you do not have access to a statistician, and financial resources are limited, consider downloading and learning R (<https://www.r-project.org/>). This is a free full-service statistical package that can be used to do just about anything you’d need to do statistically, and because it’s open source, there are new code and new macro programs being developed and posted online all the time. This has led R to rapidly move beyond SAS, SPSS, and other programs as perhaps the most versatile statistical package available. For a stand-alone power calculator, the best on the market is probably PASS (Power Analysis and Sample Size from NCSS) with calculations available for more than 150 different study design types. If your study analyses tend to be pretty straightforward, some of the free programs may be sufficient for your power needs.

Take-Home Message

- Statistical power is a commonly misunderstood and sometimes abused theoretical concept.
- While being theoretical, it has practical benefit.
- Conducting a needlessly large clinical research project is inefficient and may waste resources that could be used to further our knowledge in other ways.
- Conversely, conducting an underpowered study is a waste of time to the participants and the researchers and may violate the responsibility to do no harm as the study intervention may not be risk-free.
- All studies, especially those in which invasive interventions are being used, should always, ethically, be adequately powered.
- Power is determined by a balance of the sample size, critical p-value, variability and/or frequency, and effect size.
- Adjusting any of these characteristics will modify the estimate of statistical power, though usually sample size is the most easily modifiable in calculating power.
- Power should always be calculated a priori, though a post hoc power estimate may also be useful if the study data differs substantially from what was estimated prior to conducting the research.
- With today's powerful computing capabilities, calculating power is easier than ever, though appropriate inputs are vital.

References

1. Adcock CJ. Sample size determination: a review. *Statistician*. 1997;46(2):261–83.
2. Blackless M, Charuvastra A, Derryck A, Fausto-Sterling A, Lauzanne K, Lee E. How sexually dimorphic are we? Review and synthesis. *Am J Hum Biol*. 2000;12:151–66.
3. Carley S, Dosman S, Jones SR, Harrison M. Simple nomograms to calculate sample size in diagnostic studies. *Emerg Med J*. 2005;22:180–1.
4. Jones SR, Carley S, Harrison M. An introduction to power and sample size estimation. *Emerg Med J*. 2003;20:453–8.
5. Mcglothlin AE, Lewis RJ. Minimal clinically important difference defining what really matters to patients. *JAMA*. 2014;312:1342–3.
6. “Power.” Merriam-Webster.com. 2018. <https://www.merriam-webster.com> (1 July 2018).
7. Saito Y, Sozu T, Hamada H, Yoshimura I. Effective number of subjects and number of raters for inter-rater reliability studies. *Stat Med*. 2006;25:1547–60.
8. Sim J, Wright CC. The Kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther*. 2005;85(3):257–68.
9. Theodore BR. Methodological problems associated with the present conceptualization of the minimum clinically important difference and substantial clinical benefit. *Spine J*. 2010;10:507–9. <https://doi.org/10.1016/j.spinee.2010.04.003>.
10. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med*. 1998;17:101–10.

Stephen Lyman, Naomi Roselaar,
and Chisa Hidaka

21.1 Introduction

As you begin to analyze your data, think about how best to present it. Whether in a paper, a poster, or at a laboratory meeting, outstanding tables and figures are essential to ensure readers and viewers understand your research. In this chapter, we review how the hierarchy of human graphical perception ability elucidated by Cleveland and McGill in 1985 [1] can be used to achieve two major principles of excellent data visualization, put forth by information design pioneer Edward Tufte in his influential work, *The Visual Display of Quantitative Information* [2]: Graphical Excellence and Visual Integrity.

Tufte has stated, “Graphical elegance is often found in simplicity of design and complexity of data” [2]. In the concepts described below, and examples that follow, we’ll discuss how to achieve this goal.

Graphs and charts are able to convey complex information more efficiently than tables. However, tables are necessary when you want the reader to be able to look up individual data values. This chapter also offers guidelines for tables that are accurate and easy to understand.

Charles Minard’s map (Fig. 21.1) tells a rich, data-filled story in a single coherent visual representation, by combining geographical informa-

tion with a visual representation of the location and extent of the loss of soldiers as they advanced to and retreated from Moscow. At a glance, the huge loss is apparent, as the lines representing the number of soldiers thins during the advance, represented in gray, and dwindle down even further on the retreat, represented in black. The patterns the lines make across the page make it obvious that the path of the soldiers crossed rivers and included a variety of elevations. On a closer look, many details are available, including the names of cities or towns and, crucially, the exact number of soldiers.

This is not the type of graphic often seen in medical research but worthy of remembering as an example of the level of elegance and excellence to which all researchers should aspire.

21.2 Graphical Excellence

Fact Box 21.1

In his influential work, *The Visual Display of Quantitative Information* (2001), information design pioneer Edward Tufte summarized the following eight aspects of Graphical Excellence:

1. Show the data.
2. Avoid data distortion.

S. Lyman (✉) · N. Roselaar · C. Hidaka
Hospital for Special Surgery, New York, NY, USA
e-mail: LymanS@HSS.edu

FIGURATIVE MAP of the successive losses in men of the French Army in the RUSSIAN CAMPAIGN OF 1812-1813

Drawn by Mr. Minard, Inspector-General of Bridges and Roads in retirement, Paris, 20th November, 1869. The numbers of men present are represented by the widths of the colored zones in a rate of one millimeter for ten thousand men. The numbers of men present in the various parts of the army since 28 October, in order to facilitate the judgment of the eye regarding the diminution of the army, I supposed that the corps under Prince Adolphe and under Marshal Davoust, who were sent to Minsk and Mohlow and who rejoined near Orscha and Witebsk, had always marched with the army.

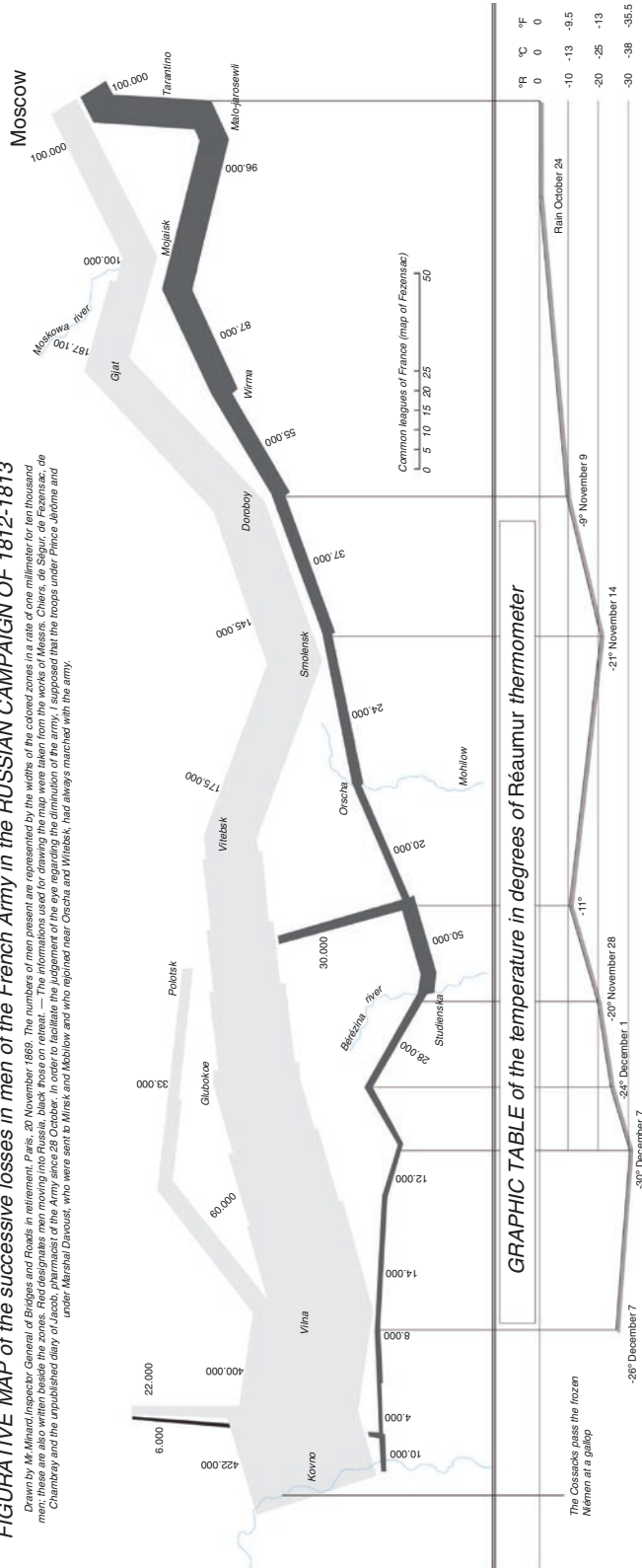


Fig. 21.1 A modern adaptation of a map by Charles Minard portraying French losses in the Russian Invasion (1812–1813) is highlighted by Tufte as an example of graphic elegance and excellence. Commons usage: https://commons.wikimedia.org/wiki/File:Minard_map_of_napoleon.png

3. Draw attention to the substance of the graph.
4. Serve a clear purpose.
5. Make large datasets coherent.
6. Reveal the data at several levels of detail.
7. Present many numbers in a small space.
8. Integrate with statistical and verbal descriptions.

Graphical excellence is achieved when the greatest amount of complex information is visually represented in a manner that allows the viewer to understand it completely, accurately, and quickly. In his influential book, *The Visual Display of Quantitative Information* [2], Tufte describes eight characteristics that define Graphical Excellence in data visualization:

1. **Show the data:** Representing as much data as possible through an economy of visual elements, without losing what is important about their patterns, is at the top of the list for Graphical Excellence. Avoid chart styles that inadvertently hide data (Example 1).
2. **Avoid data distortion:** To create an accurate graphic that shows all of the data values and patterns without distortion and follow the rules that Tufte has recommended for Visual Integrity (Sect. 21.4).
3. **Draw attention to the substance of the graph:** Focus the viewer on the substance of the graph rather than its production. Eliminate what Tufte calls “chartjunk” such as unnecessary dimensions or decorations and redundant or irrelevant information (Examples 2, 3, and 4). To present many numbers in a small space do not only eliminate “chartjunk” but maximize the “data-ink ratio.” Conveying the maximal amount of information with the greatest possible economy of visual elements is one of the cornerstones of Tufte’s prescription.
4. **Serve a clear purpose:** Excellent graphics must also serve a clear purpose. Before creating a graphic, write a short statement that summarizes what you expect viewers to understand from it. For example, “femoral tunnel malposition is the main cause for failure in anterior cruciate ligament reconstruction (ACLR)” (Example 3). Avoid descriptive statements like, “reasons for failure of ACLR”. Starting with a pointed summary statement that supports your research hypothesis ensures that you create a graphic that serves a clear purpose within your paper or presentation.
5. **Make large datasets coherent, 6. Reveal the data at several levels of detail, and 7. Present many numbers in a small space:** While these are three separate characteristics in Tufte’s definition of graphical excellence, they are complimentary and therefore described together. Once you know what you want to say with your data, use what is known about the hierarchy of human graphical perceptual capabilities (Sect. 21.3) to prioritize information visually, and create a coherent graphic that helps the viewer understand at a glance which of the multiple levels of details you reveal are the most important and/or how the details relate to each other and to your hypothesis (Example 3).
8. **Integrate with statistical and verbal descriptions:** Finally, write an appropriate legend so your figure can stand alone and the viewer can understand it without referring to the text of the manuscript. The legend should also fit into the flow of the presentation or paper without redundancy. Tufte maintains that visualizations “are paragraphs about data and should be treated as such” [2]. When your graphic is integrated with statistical and verbal descriptions and fits seamlessly into your paper or presentation like a written paragraph but can make an important, complex point in one quick glance, you have achieved graphical excellence.

21.3 Hierarchy of Human Graphical Perception Ability

Creating excellent graphics requires an understanding not only of your data but of the hierarchy of human graphical perception ability.

Cleveland and McGill [1] have shown that some visual elements are easier for people to see and understand than others.

Graphical elements in the order they are most accurately perceived:

1. Position along a common scale
2. Position along identical nonaligned scales
3. Length
4. Angle and slope
5. Area
6. Volume and density and color saturation
7. Color hue

Use this hierarchy to choose the appropriate visual element to represent your data. For example, if your data can be represented as either length or area, choose length, as it is more accurately perceived (Example 3). When conveying complex information, use the hierarchy to prioritize information visually. For example, use angle or slope for the most important information while adding shading or pattern to provide a second level of information (Example 2). When using color, saturation is more accurately perceived and understood than the relationship between hues (Example 4).

21.4 Visual Integrity

When individual data points and their analyses are translated from the spreadsheet to a graph or figure, care must be taken to avoid unintentional “visual lies” or distortions. To maintain visual integrity, Tufte offers the following rules [2]:

1. The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.
2. Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity.
3. Write out explanations of the data on the graphic itself. Label important events in the data.

4. Show data variation, not design variation.
5. The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.
6. Graphics must not quote data out of context.

In practical terms, visual integrity often comes down to proper scaling and formatting, as well as avoiding “chartjunk.” Examples follow below.

Fact Box 21.2

In practical terms, visual integrity can often be achieved through:

1. Accurate scaling
2. Proper formatting
3. Avoiding “chartjunk”

21.4.1 Scaling

Proper scaling ensures that the numbers on the graphic are directly proportional to the numerical quantities represented. It allows you to show all the data, without distortion. By considering the hierarchy of human graphical perceptual capabilities, you can choose an appropriate scale(s) that makes data patterns easy to see at a glance (Example 2).

21.4.2 Formatting

To avoid graphical distortion and ambiguity, pay attention to formatting. In practical terms: use simple symbols and clear, thorough labels and avoid remote legends (Examples 3 and 4).

21.4.3 Avoiding “Chartjunk”

To focus viewers on data rather than design, avoid unnecessary or confusing elements. Color is often used with the intention of making a graphic more appealing but can undermine the

effectiveness of the graphic, if it is not used in a manner that conveys relevant information (Examples 2, 3, and 4). Three-dimensional graphics are also used for an appealing look, but they should only be used for three-dimensional data (Examples 2 and 3). Using extra dimensions for two-dimensional data not only encumbers graphics with “chartjunk” but can create distortions or optical illusions (“visual lies”). Clearly, these should be avoided.

21.4.4 Context

Finally, an excellent graphic is well integrated into relevant descriptions and explanations in the text or presentation, avoiding the quotation of data out of context (Examples 2 and 3).

21.5 Tables

Whereas graphs and charts are superior for conveying data patterns, tables are useful when you want your reader to be able to look up specific data values.

Fact Box 21.3

When creating tables, make sure to:

1. Have a descriptive title.
2. Use appropriate headings for each column and/or row.
3. Group information in a manner that provides coherence.
4. Include appropriate decimal places for each value.
5. Provide enough information so the table can stand alone.

Footnotes are a good way to provide information such as methodology or explanation of abbreviations, so that the table can stand alone, and readers can understand it without referring to the text of the paper (Example 5).

If the tables, charts, or other graphics are being prepared for a manuscript, make sure to:

1. Follow the formatting instructions of the specific journal.
2. Make the values, abbreviations, and/or terminology in the table consistent with what appears in the text.
3. Avoid redundancy. Specific values should not appear in both the text and table.
4. Use tables for details, for instance, results instead of lengthy text.

21.6 Best Practice

In conclusion, Tufte recommends the following best practice for creating excellent graphics [2]:

1. Above all else show the data.
2. Maximize the data-ink ratio.
3. Erase non-data-ink.
4. Erase redundant data-ink.
5. Revise and edit.

In this spirit, the examples below show how graphics can be revised and edited to become more effective and efficient.

Example 1: Bar graphs

Figures 21.2 and 21.3 represent the same sample data.

The dynamite plot (Fig. 21.2) is frequently used in scientific studies, but may not reveal all of the data. In Fig. 21.2, the scores for treatment A, B, C, and D are represented as means (bars) with standard deviations (I beams).

Graphical Excellence: Show all the data

A modified box and whisker (Tukey’s) plot (Fig. 21.3) allows the viewer to see all data points, and this allows the viewer to appreciate the variation in the data much more precisely than when looking at the error bars representing standard deviation (Fig. 21.2). For each treatment group, the horizontal line represents the 50th percentile, while the rectangular box extends to the 25th and 75th percentiles. The vertical lines represent the 95% confidence interval. The dots are all of the

Fig. 21.2 Patient-reported outcome scores after treatment A, B, C, or D after N weeks of treatment

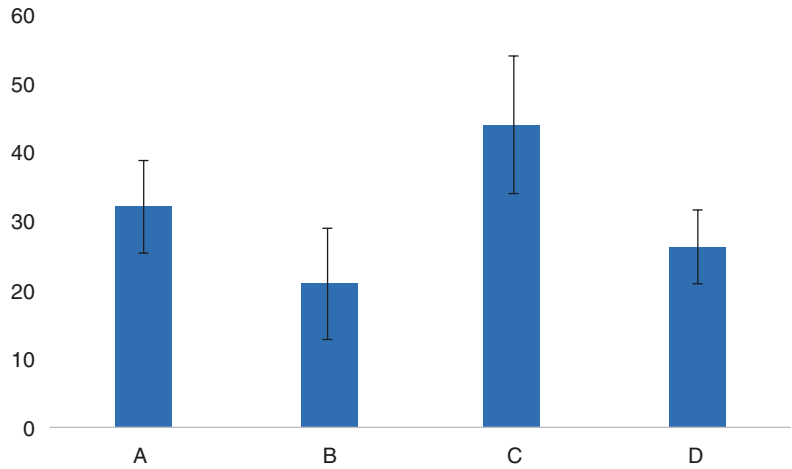
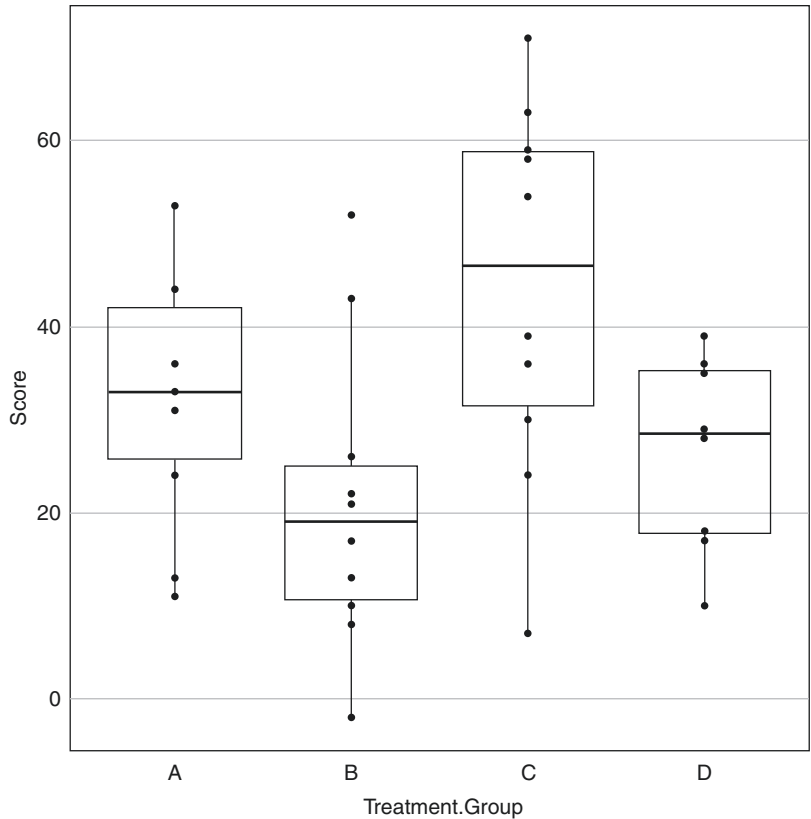


Fig. 21.3 No difference in patient outcome score after treatment for condition X with treatment A, B, C, or D. Scores were collected N weeks after treatment



data points and show that, for treatment group B, one of the points is an outlier. Comparing Figs. 21.2 and 21.3 shows that the type of plot used in Fig. 21.2 hid the presence of a negative value (below 0) in treatment group B. The comparison also reveals that treatment group D included fewer patients (data points) than the other groups.

Example 2: Line Graph

Line graphs are common in orthopedic research. This is an example showing sample data from a study where patient-reported outcome scores were assessed as a measure of treatment efficacy in a treated and (untreated) control group. Figures 21.4, 21.5, and 21.6 show the same data,

Fig. 21.4 Patient-reported outcome score after treatment for condition X

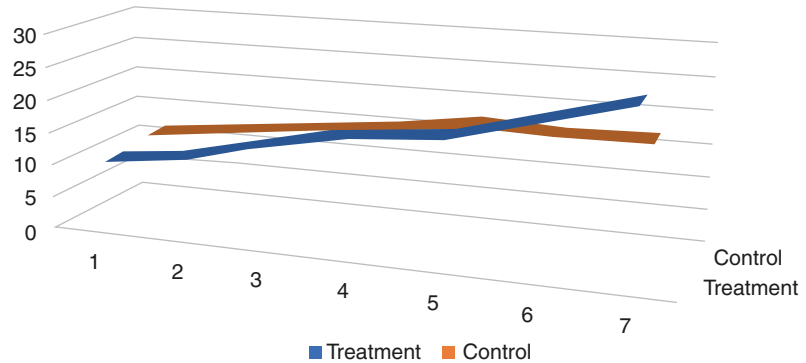
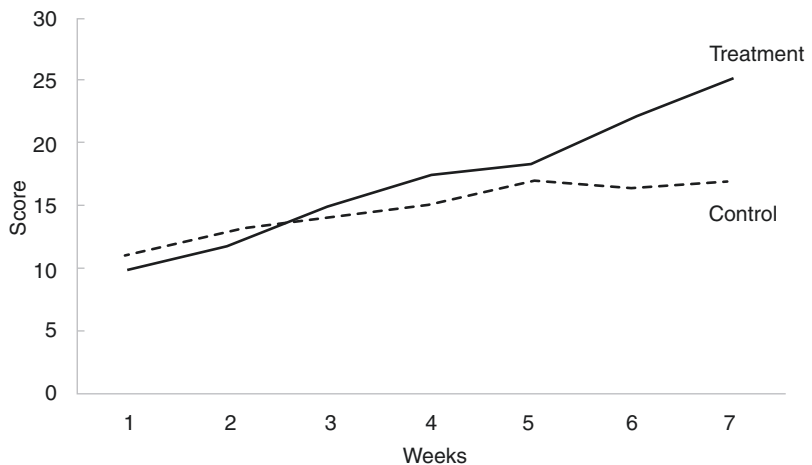


Fig. 21.5 Patient-reported outcome scores are improved with treatment. Treated ($n = 104$) and matched control ($n = 104$) patients completed the ABC survey for 7 weeks after receiving treatment X. Mean scores at each weekly time point are shown



represented with various degrees of graphical excellence and visual integrity.

Figure 21.4 is a typical chart that can be generated using Excel software and is an example of a graphic that fails to follow the rules of visual integrity and graphic excellence. Figures 21.5 and 21.6 show how the chart can be revised (within Excel) to be more effective.

Visual integrity: The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.

This data is two-dimensional, so it should be represented in a two-dimensional graph (Figs. 21.5 and 21.6). In Fig. 21.4 the third dimension not only distracts but confuses the viewer, creating an optical illusion, where scores, which are actually different in treatment and control groups, appear to overlap at several time points.

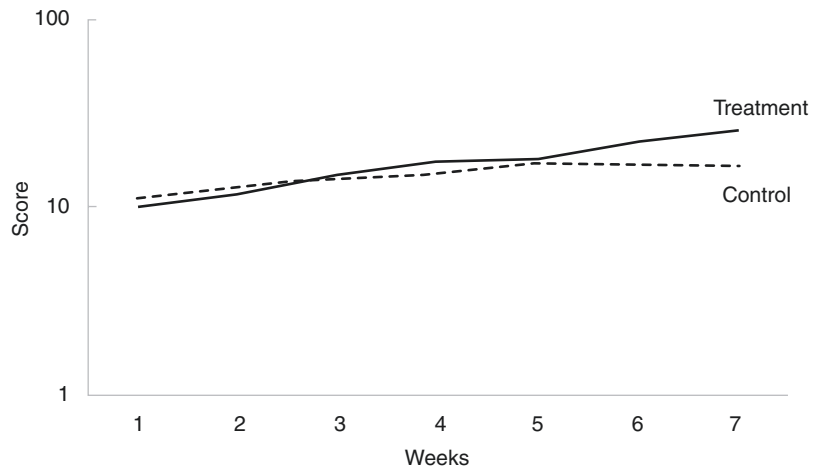
Visual integrity: Clear, detailed, and thorough labeling should be used to defeat graphical distortion and ambiguity.

Neither the X nor Y axes of this graphic are labeled (Fig. 21.4). The viewer would need to search the manuscript to understand what is represented in this graphic.

Visual integrity: Write out explanations of the data on the graphic itself. Label important events in the data.

The data lines for the treated and control groups are identified in a remote legend at the bottom of the graph (Fig. 21.4). Remote legends are not recommended, as they force the viewer to look back and forth between the legend and the data points to understand the graphic. Wherever possible, labels should appear next to the data points so viewers see the data point and what it represents in one glance (Fig. 21.5).

Fig. 21.6 No effect of treatment X on patient-reported outcomes in condition Y. Treated ($n = 104$) and matched control ($n = 104$) patients completed the ABC survey for 7 weeks after receiving treatment X. Mean scores at each weekly time point are shown



Graphical Excellence: Draw attention to the substance of the graph (not its production).

Extraneous visual elements or “chartjunk” distract the viewer. In addition to removing the extra dimension, removing gridlines, which do not improve the viewer’s ability to estimate the numerical value of each data point, improves this graph (Fig. 21.5). Placing the tick marks inside the axis labels (instead of outside) is sufficient for the eye to be able to see where the data points are located, relative to the axis scales (Fig. 21.5).

Color is also an unnecessary (and possibly distorting) element in Fig. 21.4.

Hierarchy of graphical perception: Line over color.

Color is far down the list in the hierarchy of human graphical perception, so representing information through the use of color should be avoided, if the information can be conveyed in black and white. Red and blue are particularly poor choices as people with the most common form of color blindness are often unable to distinguish red, purple, and blue.

Using a solid and dashed lines not only avoids unnecessary color, it adds information visually. A dashed line is harder to see than a solid line, so viewers will immediately understand that it is the less important line, even before reading the label, confirming that it represents the (untreated) control group (Fig. 21.5).

Graphical Excellence: Integrate with statistical and verbal descriptions.

Visual integrity: Graphics must not quote data out of context.

Figure legend 21.4, while descriptive, does not provide sufficient information for the viewer to understand the graph without referring to the manuscript. Figure 21.5 legend interprets the data and provides enough information to understand the data without additional explanation.

Hierarchy of graphical perception: Angle and slope.

With the graph in two dimensions, it is easy to see that the line representing the scores of treated patients banks at around 45° (Fig. 21.5). Extreme angles are difficult to perceive, so, where possible, choose a scale that results in lines that bank around 45°.

Visual integrity: The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.

Figures 21.5 and 21.6 show the importance of the scale. While the data are the same, the interpretation is different, based on the scale. In Fig. 21.5, the score (represented on the y-axis) is based on a 30-point survey where the differences between treated and control groups were clinically meaningful. In Fig. 21.6, the score is based on a 100-point survey where the differences between treated and control groups were clinically not significant.

Hierarchy of graphical perception: Angle and slope.

Fig. 21.7 Causes of failure after anterior cruciate ligament (ACL) reconstruction

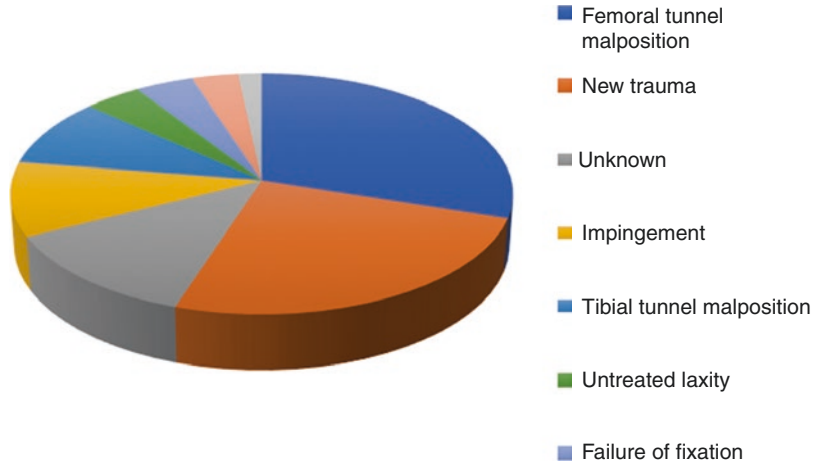
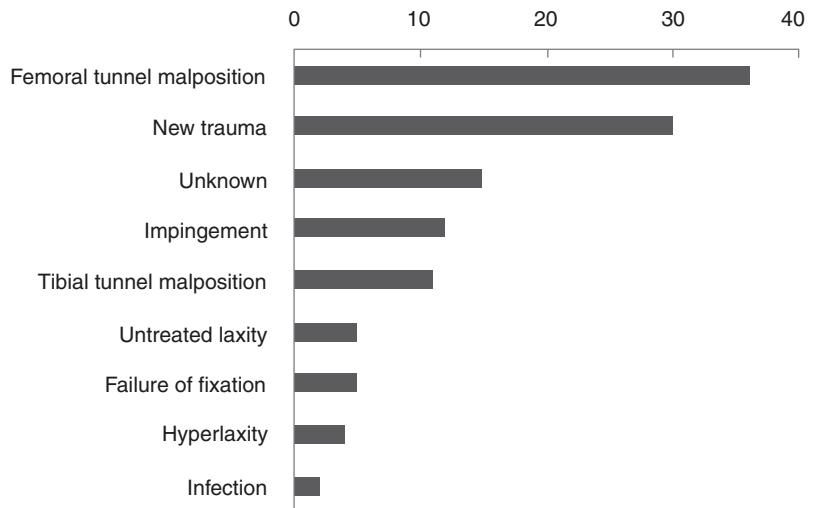


Fig. 21.8 Causes of failure after anterior cruciate ligament reconstruction



Where possible, choose a scale resulting in lines that bank around 45°, avoiding extreme angles, which are difficult to perceive.

Example 3: Pie Charts

Pie charts are often used to show how different elements account for a specific portion of the whole. However, based on the hierarchy of human graphical perception, area is perceived rather inaccurately, so alternatives are preferred, when possible.

Figures 21.7, 21.8, and 21.9 are based on data, which appear in a table format in the original publication by Trojani et al. [3].

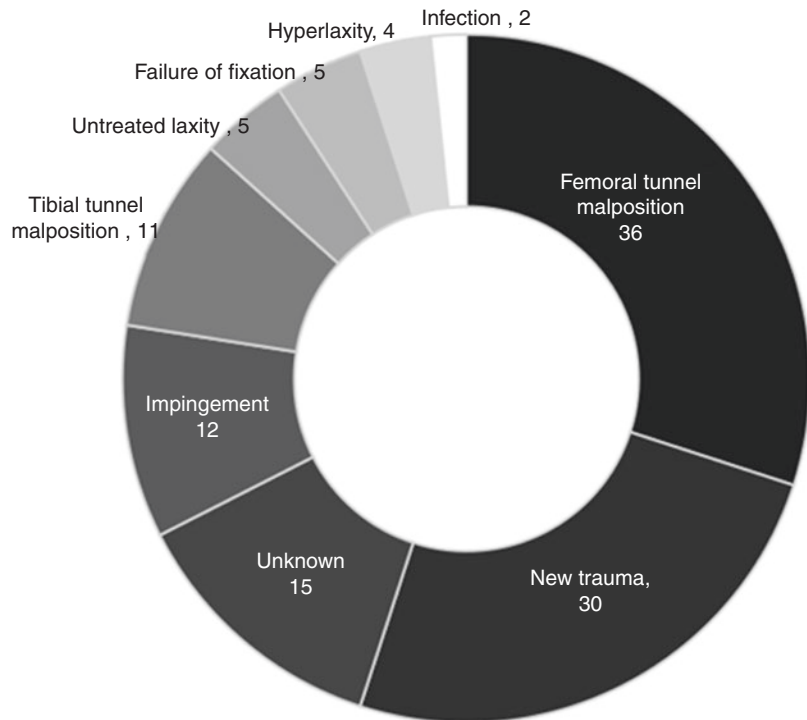
Visual integrity: The number of information-carrying (variable) dimensions

depicted should not exceed the number of dimensions in the data.

Making the pie chart in Fig. 21.7 three-dimensional adds an unnecessary dimension (“chartjunk”) as the thickness of each wedge does not convey any information. Furthermore, the use of a third dimension creates a “visual lie” (unintentional but deceptive optical illusion) in which the segments of the pie that are at the bottom of the pie appear bigger than those at the top, because the perspective used to represent the third dimension makes the thickness of the disk more apparent at the bottom of the graphic.

Visual integrity: Show data variation, not design variation.

Fig. 21.9 Femoral tunnel malposition is the main cause for failure in anterior cruciate ligament reconstruction (ACLR). The causes of failure of ACLR occurring between 1994 and 2005 in ten French orthopedic centers are shown. For each type of failure, the number of cases is indicated



The colors of the pie wedges are arbitrary adding meaningless variation that distracts, rather than informs. The colors may even imply a “visual lie,” suggesting that wedges of related colors may represent groups that are similar for some reason.

Visual integrity: Write out explanations of the data on the graphic itself. Label important events in the data.

The remote legend (Fig. 21.7) forces the viewer to look back and forth between the legend and the data (pie wedges) to understand the pie chart. Wherever possible, labels should appear next to the data points so viewers see the data point and what it represents in one glance (Figs. 21.8 and 21.9).

Hierarchy of graphical perception: Lines are perceived more accurately than areas.

Converting the information into a bar graph (Fig. 21.8) conveys the information more efficiently because length is perceived more accurately than area. However, while this type of graph makes the number of cases of each type of failure much easier to understand visually, it makes the proportion that each type of failure makes up more difficult to discern.

Graphical Excellence: Reveal the data at several levels of detail.

The circular bar chart (Fig. 21.9) shows the proportion of each type of ACLR failure, like the pie chart, but uses curved lines, rather than wedges to represent each type of failure. Instead of using arbitrary colors, a gray scale is used to accentuate visually that each adjacent arc is longer than the next so that the types of failure are arranged in order of frequency. The labels are placed within or beside each arc, avoiding remote legends. The labels are also followed by the number of cases of that type of ACLR failure, revealing another level of detail of the data.

Graphical Excellence: Integrate with statistical and verbal descriptions. Serve a clear purpose.

Visual Integrity: Graphics must not quote data out of context.

Figure legends 21.7 and 21.8 fall short of telling the viewer the point of the graphic. Figure 21.9 legend includes the author’s interpretation of the data and also provides important contextual information.

Example 4: Maps

Maps are very useful when comparing disease prevalence, treatment costs, or other phenomena that vary depending on geographical location. Figures 21.10, 21.11, 21.12, and 21.13 show data on obesity in the USA in 2016 [4].

Graphical Excellence: Make large datasets coherent.

The bar graph (Fig. 21.10) makes the proportion of each state’s population who is obese very clear, but it’s difficult to appreciate any information about relationships among states in an alphabetical list. A map is a visually efficient way to make geographic information easier to see. Figure 21.11 is a map where it is easy to see that obesity affects a greater proportion of adults in the southeast part of the USA and a lower proportion on the East and West coasts and that Colorado, Hawaii, Massachusetts, and the District of Columbia have the lowest proportion of obese adults.

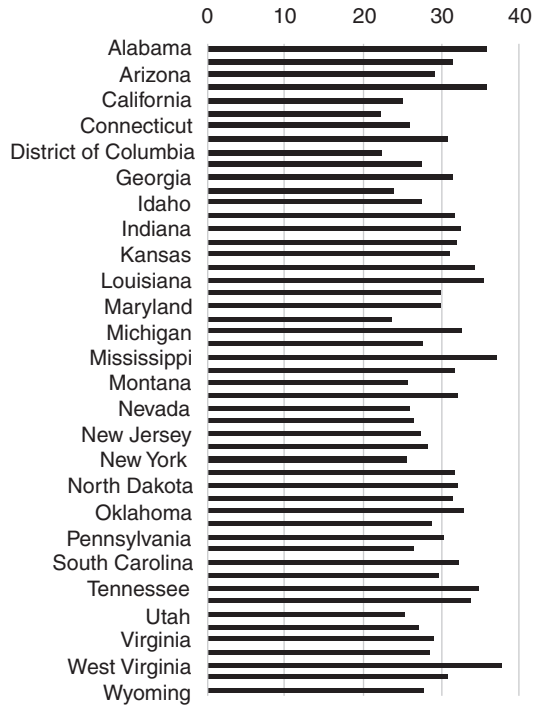


Fig. 21.10 Proportion of adults who are obese (body mass index over 30) by state in 2016

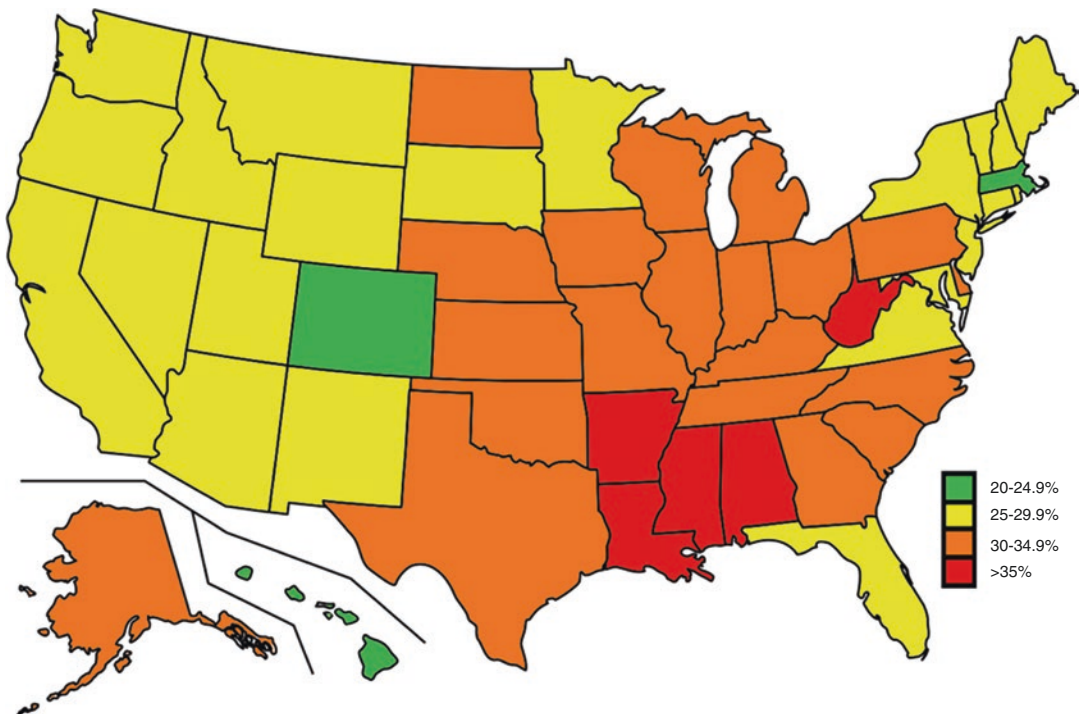


Fig. 21.11 Percentage of adults who are obese (body mass index over 30) by state in 2016

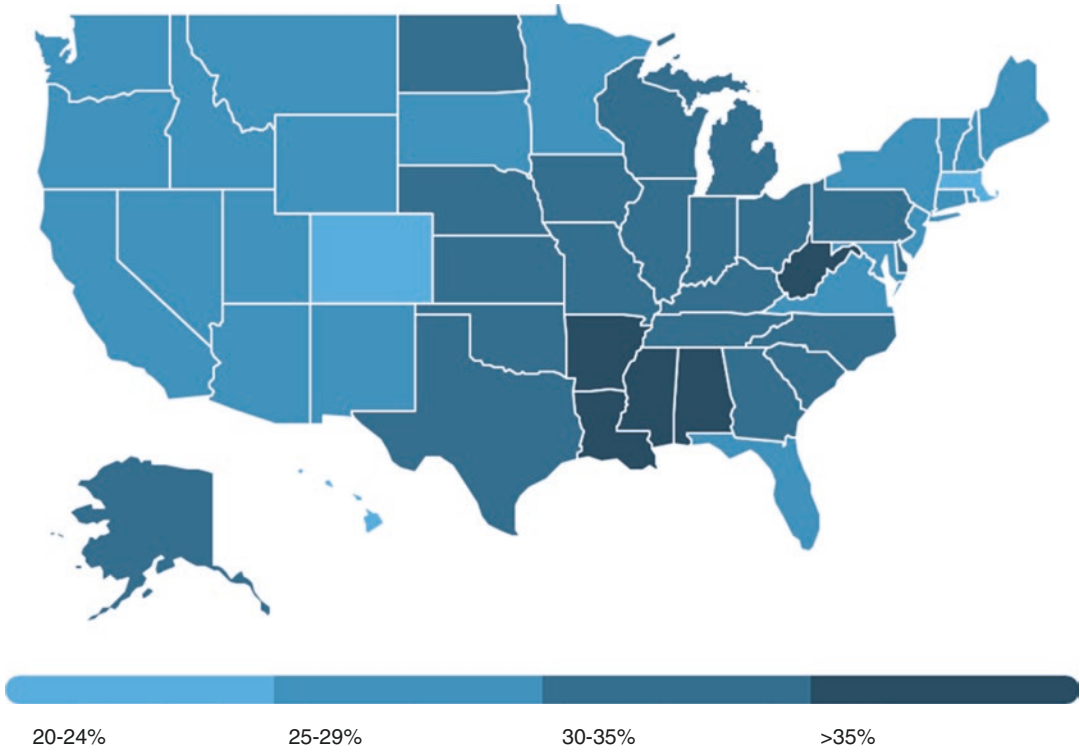


Fig. 21.12 Percentage of adults, who are obese (body mass index over 30) by state in 2016

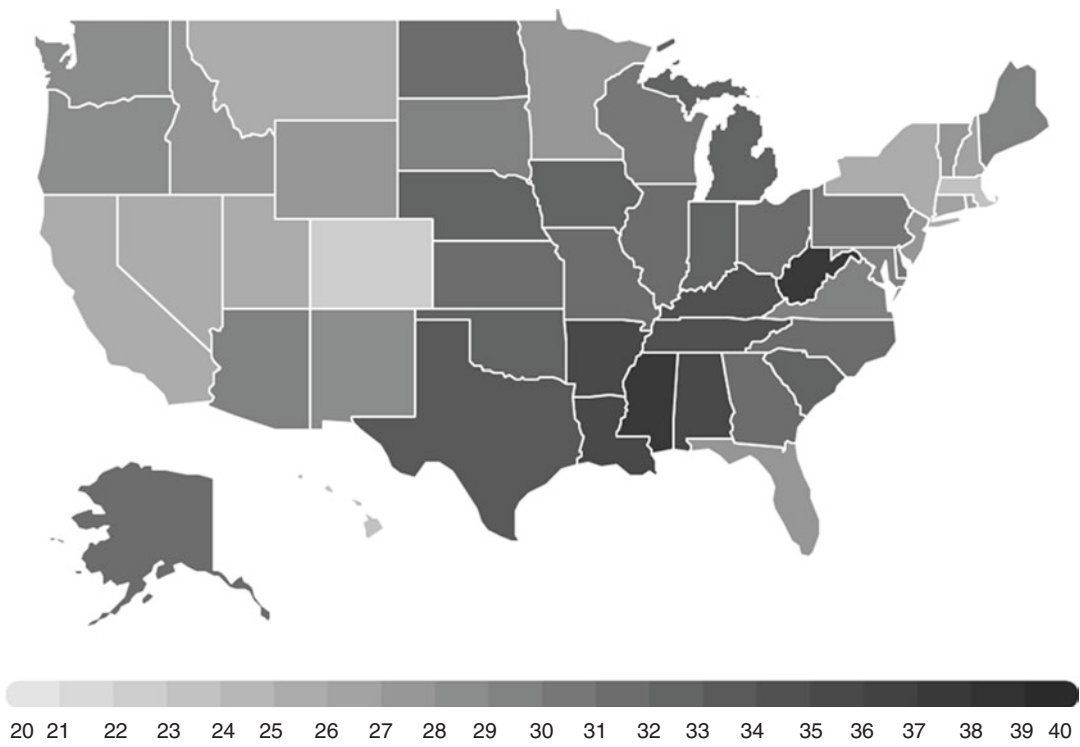


Fig. 21.13 Percentage of adults, who are obese (body mass index over 30) by state in 2016

Hierarchy of graphical perception: Color.

The use of color in Fig. 21.11 is confusing. Although the “color wheel” has been used so that relationships between colors have a direct relationship to the data, this relationship is difficult to discern. The fact that red indicates a proportion much higher than green does make sense based on the “color wheel” but is not obvious without having to look at the remote legend.

Use of color should be limited to graphics where the color itself conveys important information, as color is low in the hierarchy of graphical perception. Assessing relationships between colors accurately is difficult even for the normally sighted and impossible for the significant proportion of people who are color blind.

Hierarchy of graphical perception: Color.

In Fig. 21.12, color has been replaced by hue. Hues are less accurately perceived than color, when considering the hierarchy of human graphical perception. However, the relationship between hue and number is easily understood. In Fig. 21.12, lighter represents lower and darker represents higher numerical values—a relationship that is easy to see.

Visual economy.

In Fig. 21.12, the two-letter abbreviation for each state has been removed as most people can

easily identify the state in a map without the label, making the labels redundant and a distraction from the information.

Visual integrity: The representation of numbers, as physically measured on the surface of the graphic itself, should be directly proportional to the numerical quantities represented.

Using a gray scale allows the appropriate representation of the data as a continuous variable. Figures 21.11 and 21.12 represented the prevalence of obesity as if they occurred as categories (20–24%, 25–29%, 30–34%, or >35%), but these categories were arbitrary. Whether a state has a proportion of adults in a specific category (20–24%, 25–29%, 30–34%, or >35%) has no inherent meaning. Although a graduation from white to a color (e.g., teal, in Fig. 21.12) could also represent values in continuity, eliminating the unnecessary color removes a distraction, so that the viewer can focus on the data.

Example 5: Tables

These tables represent sample data comparing two surveys for the assessment of hand function.

Table 21.1 title is descriptive but redundant, including information that is repeated in the table itself. Information to help the reader understand the table should appear in a footnote (Table 21.2).

Table 21.1 Responsiveness of Survey A and Survey B, showing the means, and 95% confidence intervals for the scores at injection day, Day 30, and the change from injection day to Day 30

Survey	Time point	N	Mean	p-value	Effect size	Standard response mean
A	Injection day	109	42 (38, 46)	<0.001	1.5	1.2
	Day 30	88	72 (68, 76)			
	Change	87	30 (25, 35)			
B	Injection	109	20 (16, 23)	<0.001	−0.5	−0.7
	Day 30	89	12 (9, 14)			
	Change	88	−9 (−11, −6)			

The *p*-values indicate the significance of the difference between scores at Day 30 and injection day. Effect sizes and standard response means for each instrument are also displayed

Table 21.2 Responsiveness of Survey A is greater than Survey B in patients with Dupuytren’s disease

	Survey A (N = 87)	Survey B (N = 88)
Mean change in score (95% CI) ^a	30 (25, 35)	−9 (−11, −6)
Effect size	1.5	−0.5
Standard response mean	1.2	−0.7

^aMean change in score between Day 0 (treatment with a collagenase injection) and 30 days after treatment for Survey A (87 patients) and Survey B (88 patients) are reported. The 30-day mean change in score was significant for each survey (both *p* < 0.001).

Table 21.1 is also confusing, showing patient number (N) for different time points, when the only important information is the change in mean survey scores between the day of treatment and 30 days later (change). Also, even though the effect size and standard response mean are the most important information, they are far away from the name of the survey instrument whose effectiveness they report.

Table 21.2 title summarizes the analysis and provides an interpretation to help readers understand the data. A footnote provides methodological information required to understand the information, so that the table stands alone and readers do not need to refer to anything else (such as the rest of the manuscript or poster).

Tables should be used when it is important for the viewer to see the specific numbers that are shown. Some of the values in Table 21.1 have been removed so that Table 21.2 includes only the essential numbers to convey the information from the analysis. Streamlining the information also makes it possible to remove unnecessary boxes and lines.

For the information in Tables 21.1 and 21.2, a table is better suited than a graph. A graph would show positive data points for Survey A versus negative ones for Survey B, and this would create a “visual lie,” where the fact numbers above or below zero appear important, whereas due to differences in the way the two surveys are scored, it

is the absolute magnitude of the value that is relevant.

Take-Home Message

- The goal is not to make tables and figures look nice but to make sure that viewers understand your research accurately and efficiently. Just as understanding grammar and syntax make for clear writing, understanding the hierarchy of human graphical perception ability is essential for creating graphs and charts that make complex information easy to understand.
- The appropriate use of figures and tables, created with Graphical Excellence and Visual Integrity, is as important as well-written text in the dissemination of research.

References

1. Cleveland WS, McGill R. Graphical perception: theory, experimentation, and application to the development of graphical methods. *J Am Stat Assoc.* 1984;79(387):531–54.
2. Tufte ER. *The visual display of quantitative information.* 2nd ed. Cheshire: Graphics Press; 2001.
3. Trojani C, Sbihi A, Djian P, Potel JF, Hulet C, Jouve F, Bussière C, Ehkirch FP, Burdin G, Dubrana F, Beaufils P. Causes for failure of ACL reconstruction and influence of meniscectomies after revision. *Knee Surg Sports Traumatol Arthrosc.* 2011;19(2):196–201.
4. The state of obesity, trust for America’s Health and Robert Wood Johnson Foundation. <https://stateofobesity.org/adult-obesity/>.

Part IV

Basic Toolbox for the Young Clinical Researcher



How to Prepare an Abstract

22

Elmar Herbst, Brian Forsythe, Avinesh Agarwalla,
and Sebastian Kopf

22.1 Introduction

Orthopaedic research encompasses technical notes, case reports, systematic reviews, meta-analyses, retrospective studies, and prospective studies. While there is a continuum of strength associated with each type of research, the commonality in the dissemination of all orthopaedic research is the abstract—which serves as a snapshot of the work that was completed [8, 9, 13, 14]. Abstracts are submitted for review in conferences, and if accepted, it is indexed as part of the proceeding. When fellow researchers then read the published paper, the manuscript is rarely examined in its entirety on the first pass. Rather, researchers will carefully review the abstract to understand the study’s purpose, methodology, results, and conclusion. From this brief overview, readers determine if the study fits their needs. If

the study fits, then they will generally examine the introduction and discussion of the paper, and only readers who have a particular interest in the topic will read the entirety of the paper. In many ways, the abstract functions as the most important part of the manuscript because it “sells” the publication to readers. If readers cannot understand the study during a brief overview, the manuscript is unlikely to undergo further critical review regardless of if the study has an impeccable design or impactful results.

22.2 Types of Abstracts

There are two general types of abstracts: descriptive and informative. A descriptive abstract is approximately 100 words detailing the purpose, goals, and methods of the article. In a descriptive abstract, a description of the study’s results is not provided; thus, it is necessary to read the article in its entirety to view the results. The purpose of this type of abstract is to introduce the subject to readers, who must then read the entire paper to learn about the results and conclusions of the investigation (Example 1). Informative abstracts, on the other hand, are approximately 350 words and serve as an overview of the entire paper detailing the purpose, background, methods, results, and conclusions. Informative abstracts provide data on the content of the work and highlight salient points of the study in its entirety so

E. Herbst
Department of Orthopaedic Sports Medicine, TU
Munich, Munich, Germany

B. Forsythe · A. Agarwalla
Midwest Orthopaedics at Rush, Rush University
Medical Center, Chicago, IL, USA
e-mail: brian.forsythe@rushortho.com

S. Kopf (✉)
Center for Orthopaedics and Traumatology, Hospital
Brandenburg, Medical School Theodor Fontane,
Brandenburg an der Havel, Germany
e-mail: s.kopf@klinikum-brandenburg.de

that readers can understand the study and its implications in its entirety without having to read the complete text (Example 2). These types of abstracts generally follow a format set forth by the publishing journal.

It should be noted, however, that abstracts that are submitted to conferences are generally longer (approximately one page in its entirety) and a figure and/or table can be provided. Abstracts that are submitted to conferences follow the same format as an abstract that describes a manuscript, but generally more emphasis can be placed on presenting the results or discussing the findings of the investigation.

Example 1: Descriptive Abstract [6]

Posterior cruciate ligament (PCL) reconstruction generally uses an Achilles tendon allograft, although recently, quadriceps tendon has been used an alternative option due to its size and high bone density. In this investigation, we compared the biomechanical strength of quadriceps tendon versus an Achilles tendon allograft during PCL reconstruction. Thirty fresh-frozen cadaveric knees were assigned to one of the three groups: (1) intact PCL, (2) PCL reconstruction with quadriceps tendon allograft, or (3) PCL reconstruction with Achilles tendon allograft. Posterior tibial translation was measured at neutral and 20° external rotation, and then each specimen underwent a preload, cyclic loading protocol of 250 cycles, and load to failure.

Example 2: Informative Abstract [6]

Background: Previous investigations of posterior cruciate ligament (PCL) reconstruction suggest that normal stability is not restored in many patients. The Achilles tendon allograft is frequently used, although recently, the quadriceps tendon has been utilized due to its size and high bone density.

Purpose: The purpose of this investigation was to compare the biomechanical strength of a quadriceps versus an Achilles allograft during PCL reconstruction. We hypothesize that quadriceps allografts have comparable mechanical properties to those of Achilles allografts.

Methods: Thirty fresh-frozen cadaveric knees were assigned to one of the three groups: (1) intact PCL, (2) reconstruction with quadriceps

tendon allograft, or (3) reconstruction with Achilles tendon allograft. Posterior tibial translation was measured at neutral and 20° external rotation, and then each specimen underwent a preload, cyclic loading protocol of 250 cycles, and, lastly, load to failure.

Results: The intact specimens achieved the greatest failure load compared to both reconstructions (2048 ± 969 N, $p = 0.001$). Quadriceps tendon allografts had a higher maximum force during failure testing than the Achilles allograft (2017, 1837 N, respectively, $p = 0.007$). No significant differences were noted between quadriceps and Achilles allograft for differences in displacement of the graft, creep deformation, or stiffness. Construct stiffness during failure testing was greatest in the intact group (169 ± 9 N/mm, $p = 0.0005$) compared with the Achilles (56 ± 13 N/mm) and quadriceps (47 ± 3 N/mm) groups.

Conclusion: Quadriceps and Achilles tendon allografts had similar biomechanical properties when used for a PCL reconstruction, but both were inferior to the native PCL. However, quadriceps tendon allografts displayed a stronger construct with failure load and stiffness in comparison to Achilles tendon allografts.

Clinical relevance: The quadriceps tendon is a viable graft option in PCL reconstruction as it exhibits a greater maximum force but is otherwise comparable to the Achilles allograft. The findings of this investigation expand allograft availability in PCL reconstruction.

Apart from descriptive and informative abstracts, structured abstracts must be distinguished from unstructured abstracts. Structured abstracts follow a clear structure as mentioned below (e.g. background, methods, results, and conclusion; Example 2) [12]. In contrast, unstructured abstracts do not follow distinct paragraphs but represent a running paragraph as shown in the Example 3.

Example 3: Unstructured Abstract of a Narrative Review [2]

Meniscectomy is one of the most popular orthopaedic procedures, but long-term results are not entirely satisfactory, and the concept of meniscal preservation has therefore progressed over the years. However, the meniscectomy rate remains too high even though robust scientific publica-

tions indicate the value of meniscal repair or non-removal in traumatic tears and nonoperative treatment rather than meniscectomy in degenerative meniscal lesions. In traumatic tears, the first-line choice is repair or non-removal. Longitudinal vertical tears are a proper indication for repair, especially in the red-white or red-red zones. Success rate is high and cartilage preservation has been proven. Non-removal can be discussed for stable asymptomatic lateral meniscal tears in conjunction with anterior cruciate ligament (ACL) reconstruction. Extended indications are now recommended for some specific conditions: horizontal cleavage tears in young athletes, hidden posterior capsulo-meniscal tears in ACL injuries, radial tears, and root tears. Degenerative meniscal lesions are very common findings which can be considered as an early stage of osteoarthritis in middle-aged patients. Recent randomized studies found that arthroscopic partial meniscectomy (APM) has no superiority over nonoperative treatment. Thus, nonoperative treatment should be the first-line choice, and APM should be considered in case of failure: 3 months has been accepted as a threshold in the ESSKA Meniscus Consensus Project presented in 2016. Earlier indications may be proposed in cases with considerable mechanical symptoms. The main message remains: save the meniscus!

Commonly, unstructured abstracts belong to narrative reviews of current literature where a structured design would not be appropriate because of the lacking Results or Methods sections. However, some basic research journals ask also for unstructured abstracts (Example 4). Regardless of the type of abstract, each abstract should provide a brief background or introduction to the topic, a summary of the methods, and the results followed by a concluding remark as described more in detail in the following sections.

Example 4: Unstructured Abstract of an Experimental Study [3]

Ligament and tendon repair is an important topic in orthopaedic tissue engineering; however, the cell source for tissue regeneration has been a controversial issue. Until now, scientists have been split between the use of primary ligament fibroblasts and bone marrow-derived mesenchymal

stem cells (MSCs). The objective of this study was to show that a coculture of anterior cruciate ligament (ACL) cells and MSCs has a beneficial effect on ligament regeneration that is not observed when utilizing either cell source independently. Autologous ACL cells (ACLc) and MSCs were isolated from Yorkshire pigs, expanded *in vitro*, and cultured in multiwell plates in varying %ACLc/%MSC ratios (100/0 75/25, 50/50, 25/75, and 0/100) for 2 and 4 weeks. Quantitative mRNA expression analysis and immunofluorescent staining for ligament markers collagen type I (collagen-I), collagen type III (collagen-III), and tenascin C were performed. We show that collagen-I and tenascin C expression is significantly enhanced over time in 50/50 cocultures of ACLc and MSCs ($p \leq 0.03$), but not in other groups. In addition, collagen-III expression was significantly greater in MSC-only cultures ($p \leq 0.03$), but the collagen-I-to-collagen-III ratio in 50% coculture was closest to native ligament levels. Finally, tenascin C expression at 4 weeks was significantly higher ($p \leq 0.02$) in ACLc and 50% coculture groups compared to all others. Immunofluorescent staining results support our mRNA expression data. Overall, 50/50 cocultures had the highest collagen-I and tenascin C expression and the highest collagen-I-to-collagen-III ratio. Thus, we conclude that using a 50% coculture of ACLc and MSCs, instead of either cell population alone, may better maintain or even enhance ligament marker expression and improve healing.

22.3 Components of an Abstract

Depending on the journal, an abstract can be written as a free-flowing paragraph or each component of the abstract must be separated and formatted into a formal structure. Some journals may require additional sections that discuss the clinical relevance, limitations, or a description of the study design.

22.3.1 Background

The background of an abstract is generally one to two sentences that present the problem that the investigator aims to address. In this short space,

the questions that need to be answered are “why is this study important?” and “what is the impact of this study?”. A good statement in background section tells the reader what is already known about the topic and what needs to be investigated. A common mistake in this section is to discuss what has already been studied about the topic but fails to discuss the gap that remains in the literature on that topic. A statement of what needs to be further investigated does not need to be explicitly stated; rather the reader should be able to infer what further research is warranted. Examples of a good background section and the most common pitfall are provided in Table 22.1. The background section sets the stage for the reader to understand

Table 22.1 Examples of appropriate and insufficient background statements in an abstract with comments on the differences between the statements (bolt)

Appropriate background statements
<ul style="list-style-type: none"> • Previous investigations of posterior cruciate ligament (PCL) reconstruction suggest that normal stability is not restored in many patients. (<i>Why is the study needed?</i>) The Achilles tendon allograft is frequently used, although recently, the quadriceps tendon has been utilized due to its size and high patellar bone density [6] (<i>What is known about the subject?</i>) • During arthroscopic ACL reconstruction, transtibial drilling techniques are widely used because they simplify femoral tunnel placement and reduce surgical time (<i>What is known about the subject?</i>) However, there has been concern that this technique results in non-anatomically positioned bone tunnels, which may cause abnormal knee functionality [10]. (<i>Why is the study needed?</i>)
Insufficient background statements
<ul style="list-style-type: none"> • PCL reconstruction outcomes are variable and have not had the same success as ACL reconstructions (<i>research question not evident and not specific enough</i>). Numerous surgical techniques have been described, including open and arthroscopic tibial inlay and transtibial techniques. Currently, the most frequently utilized technique is an open transtibial technique with an Achilles tendon allograft [6] (<i>insufficient background information related to the research question</i>) • During arthroscopic ACL reconstruction, transtibial drilling techniques are widely used because they simplify femoral tunnel placement and reduce trauma and surgical time by employing a single-incision approach. Thus, the transtibial approach was seen a major innovation and quickly became a popular technique for arthroscopic ACL reconstruction [10] (<i>research question not provided; the reviewer is not informed why the study is needed</i>)

the research gap that this study aims to fill without the reader having to conduct an extensive literature search. It’s important to keep the background of the abstract short. While an extensive background allows the reader to be well-informed of previously established knowledge regarding the topic, it prohibits a more detailed discussion of the present investigation [1, 4, 5].

22.3.2 Purpose

The purpose is another one to two sentences illustrating the problem statement of the investigation. This portion of the abstract needs to address the question “What is this study investigating?” The statement needs to be a concise explanation of what the study specifically aims to investigate. Within this section, the scope should be identifiable—whether the study is addressing a specific problem or an overarching generic issue. In other words, the scope should identify the applicability of the investigation. For example, “The purpose of this investigation is to compare the biomechanical strength of a quadriceps tendon versus an Achilles tendon allograft during a transtibial PCL reconstruction” [6]. In this example, the purpose of the investigation (biomechanical strength of different grafts) and the scope of the investigation (PCL reconstruction using a transtibial approach) are clearly stated such that the reader understands what the investigators plan to conduct. It is common to see abstracts combine the purpose with the background into a single section. Together, the background and purpose serve to tell the reader what has already been established in the literature and what this current study aims to investigate.

Fact Box 22.1: Background and Purpose

- Highlight the rationale and importance of the study
- Do not summarize what has already been studied
- The purpose should be brief and follows the background sentence
- The purpose should be a clear statement on what the study is investigating

22.3.3 Methods

The Methods portion is generally the second longest section of the abstract that needs to address the following questions: “How was the study conducted?” “Who/what was included in the study?” “What were the investigational arms?” “How many subjects were included?” “What was measured?” “When were measurements taken?” “How was the study designed?” After reading this section, the reader should have a clear understanding of how the study was carried out by the investigators. In general, statistical analyses that were employed in the investigation are not included in the methods section unless there was a special statistical test performed (i.e. Bonferroni-adjusted *t*-test) that assisted in interpreting the results. The platform used to conduct the statistical analyses is not mentioned in the abstract, and those details are discussed within the manuscript itself.

While this section is designed to help the reader understand how the study was conducted, it is easy to include extraneous information that

takes up much-needed space in the abstract and may confuse the reader. It is important to carefully detail a description of the study design in a succinct manner that allows the reader to generally understand how the study was completed. Gaps in this section can confuse the reader and cause them to question the credibility of the study’s results. An example of an appropriate and improper discussion of methodology is provided in Table 22.2. With a proper discussion of methods, the reader is able to understand the salient details of how the study was carried out. On the other hand, an improper description of the methods includes immense detail that prevents the reader from quickly and effectively understanding the study. In the provided example, much of the detail provided is appropriate for inclusion within the manuscript itself as it provides the reader a detailed and reproducible description of the study. In the abstract, only a brief description describing what was conducted should be included, while details allowing the reader to reproduce the study should be included within the manuscript itself [1, 5].

Table 22.2 Appropriate and improper description of a study’s methods in an abstract

Improper description of methodology	Proper description of methodology
34 patients scheduled to undergo a primary ACL reconstruction were prospectively evaluated in this investigation. The dial test was performed under anaesthesia on both the affected and unaffected knees at 30° and 90° of flexion with a goniometer by two examiners. Intraoperatively, PLC gaps were evaluated prior to reconstruction and immediately following. Postoperatively, the dial test was again administered at 30° and 90° on both knees by the same two examiners [7] (<i>From this description, it is not clear whether the study was prospective or retrospective, how the authors collected and quantified the data, and how the statistical analysis was performed</i>)	From April 2017 to May 2017, consecutive patients from a single institution scheduled to undergo a primary ACL reconstruction were prospectively evaluated. Following anaesthesia, the dial test was performed on both lower extremities at 30° and 90° of knee flexion by two examiners with the patient supine. Both examiners performed all evaluations throughout the study to promote consistency. Tibial external rotation was measured with a goniometer. Intraoperatively, with the knee at 30° of flexion and with a varus load applied to the knee, the PLC gap was measured with a calibrated nerve hook prior to reconstruction and following reconstruction. Knees with more than 14 mm of lateral tibiofemoral compartment opening are considered to have incompetent PLC. Postoperatively, the dial test was again performed on both the affected and unaffected knees at 30° and 90° of knee flexion by the same examiners using a goniometer. With an expected difference of $11 \pm 5^\circ$ of external rotation for the affected extremity on dial test before and after ACL reconstruction and with statistical power set at 0.8, the sample size needs to be at least 25 patients. The values from examiner #1 and examiner #2 were averaged to provide mean numeric values for tibial external rotation with dial test performance. Intraclass correlation (ICC) was assessed for both examiners. Two-sample paired <i>t</i> -test was used to generate 95% confidence intervals and <i>p</i> -values for each investigated condition [7] (<i>This Methods section provides information on the study design, investigation, and data analysis as well as proper statistics section</i>)

Fact Box 22.2: Methods

- Focus on the key elements of the study design to properly inform the reader
- Measurement methods must be given accurately
- Special statistical tests (i.e. Bonferroni correction) should be mentioned

22.3.4 Results

This is the point of the abstract that begins to differentiate between descriptive and informative abstracts, and it functions as the most important component of the abstract because readers are interested in learning about the findings of the investigation. The purpose of this section is to answer the question “What was found?” The results of the study should be presented clearly in as much detail as possible. The length or quality of this section should not be compromised. The

data presented in the abstract should be consistent with the findings presented in the manuscript. A recent study demonstrated that the findings presented in the abstract did not accurately reflect the manuscript in nearly 78% of cases [11]. An example of appropriate and insufficient results sections is provided in Table 22.3.

Well-written and detailed abstracts will not only describe the results but will also include numerical values, such as mean, median, standard deviation, and statistical comparisons, to fortify their statements [1]. When discussing statistical significance, only the p -value is included as a parenthetical note after listing the result. Specific measures that describe the statistical test, such as t -value or degrees of freedom, are excluded from the abstract. In general, it is best to not to describe the results vaguely by using words such as “small,” “massive,” “unlikely,” or “significant”; rather clear and direct statements should be made to describe the results. While it’s important to disseminate the results of the investigation, every single result does not need to be documented in

Table 22.3 Description of a proper and inappropriate description of the results of an investigation in an abstract

Improper description of results	Proper description of results
At 30°, there was a significantly larger dial test result in the affected knee prior to ACL reconstruction compared to after ACL reconstruction (29.6° vs. 19.0°, $p < 0.0001$) and compared with the unaffected knee (29.6° vs. 22.5°, $p < 0.0001$), but this difference was eliminated after reconstruction (14.0° vs. 13.5°, $p = 0.69$). At 90°, there was a significantly larger dial test result in the affected knee before ACL reconstruction compared with after ACL reconstruction (31.6° vs. 21.1°, $p < 0.0003$) and compared with the unaffected knee (31.6° vs. 21.9°, $p < 0.0001$); with this difference was eliminated after reconstruction (12.1° vs. 11.9°, $p = 0.3189$) [7]. <i>(Even though the results of the dial test are presented, no information on the study cohort are given. Data presentation lacks standard deviations and/or confidence intervals, making proper understanding of the results difficult. Furthermore, the authors did not present the data as described in the methods section (Table 22.2))</i>	Thirty-eight consecutive patients who underwent ACL reconstruction were prospectively evaluated in the 6-month study period between April 2017 and May 2017. The mean age of the included patients was 32.0 ± 12.6 years, with mean BMI of 26.3 ± 7.3 kg/m ² . Most patients were male (58.6%) with sports-related injuries (66.0%) and a median time between injury and surgical intervention of 31 days. The ICC between both examiners was 0.969, indicating a high reliability of the gathered measurements. At 30°, there was a significantly larger dial test result in the affected knee prior to ACL reconstruction compared to after ACL reconstruction (29.6° vs. 19.0°; 95% CI [-4.9, -6.6]; $p < 0.0001$) and compared with the unaffected knee (29.6° vs. 22.5°; 95% CI [5.8, 7.4]; $p < 0.0001$), but this difference was eliminated after reconstruction (14.0° vs. 13.5°; 95% CI [0.7, -1.2]; $p = 0.69$). At 90°, there was a significantly larger dial test result in the affected knee before ACL reconstruction compared with after ACL reconstruction (31.6° vs. 21.1°; 95% CI [-4.9, -6.9]; $p < 0.0002$) and compared with the unaffected knee (31.6° vs. 21.9°; 95% CI [-5.2, -9.4]; $p < 0.0001$); with this difference was eliminated after reconstruction (12.1° vs. 11.9°; 95% CI 1.31, -1.7]; $p = 0.41$). The PLC gap was measured at less than the critical 12 mm both pre-ACL reconstruction (6.9 mm) and post-ACL reconstruction (6.3 mm). The PLC gap decreased significantly (95% CI [-0.1, -0.7]; $p = 0.0009$) [7]. <i>(This results section follows a clear structure with demographics first, followed by inter-rater reliability data and the primary outcome variables. The authors provide confidence intervals as well as data on the aforementioned PLC gap in accordance to the methods section in Table 22.2)</i>

Fact Box 22.3: Results

- Data should be consistent with the results in the manuscript
- Include numerical values (i.e. mean, standard deviation, confidence interval)
- If the results section is too long, focus on the main findings
- Whenever the term significant is used, it should be followed by a *p*-value

the abstract as this may convolute the overall message. If the result can be easily misinterpreted, it should be excluded from the abstract but described in greater detail within the manuscript where it can be more clearly explained.

In the provided example, the results of the study were not significant at 90° and therefore do not need to be included in the abstract. However, it is important to have these results within the manuscript itself. Additionally, in this study, the PLC gap was measured to evaluate for concomitant PLC injury. Patients with concomitant PLC injury were removed from the study. While these results are important within the manuscript, they are extraneous to the abstract.

22.3.5 Conclusion

This section of the abstract contains the most concise and important take-away message from the study. This portion of the abstract is generally one to two sentences and answers the question “What are the implications of the investigation?” In addition to the primary discovery of the study, other important findings should be described as well. The goal of this section is to describe how the results of this investigation fit into the scope of previously conducted research and how the study impacts our current knowledge base. Since readers may skip directly to this section, it is the authors’ responsibility to make a concise and accurate assessment of the results of the investigation and its implications [1].

The conclusion should also incorporate a statement that discusses the clinical relevance of

the study. While it is important to document the primary discovery of a study, it is of particular importance to illustrate how that finding impacts clinical practice to improve patient care and outcomes. If the study was of a basic science nature, the authors should discuss how the conclusions will stimulate further clinical investigations or alter clinical practice. Some journals may require that the clinical relevance be separated from the conclusion and placed within a separate heading within the abstract. Examples of concluding remarks are provided in Table 22.4. Appropriate concluding remarks discuss the main finding of the investigation and how it fits into clinical practice. On the other hand, insufficient concluding remarks discuss the main finding of the investiga-

Table 22.4 Appropriate and poor examples of concluding remarks within an abstract

Appropriate concluding remarks
<ul style="list-style-type: none"> • Quadriceps and Achilles tendon allografts had similar biomechanical properties when used for a PCL reconstruction, but both were inferior to the native PCL. The quadriceps tendon is a viable graft option in PCL reconstruction as it exhibits a greater maximum force but is otherwise comparable to the Achilles allograft. The findings of this investigation expand allograft availability in PCL reconstruction [6] <i>(after a summary of the key findings, the authors provide a sentence related to the clinical relevance of the study)</i> • Incompetence of the ACL accounts for nearly 10° of tibial external rotation as evidenced by the dial test. If the dial test is positive during examination of a traumatic knee injury, an isolated ACL injury should not be excluded. Thus, findings of the dial test should thus be interpreted with caution in the setting of ACL injury [7] <i>(after a summary of the key findings, the authors provide a sentence related to the clinical relevance of the study)</i>
Insufficient concluding remarks
<ul style="list-style-type: none"> • Quadriceps and Achilles tendon allografts had similar biomechanical properties when used for a PCL reconstruction, but both were inferior to the native PCL. However, quadriceps tendon allografts displayed a stronger construct with failure load and stiffness in comparison to Achilles tendon allografts [6] <i>(in this example, the concluding sentences are somewhat contradicting each other. Furthermore, the clinical relevance is missing)</i> • Incompetence of the ACL accounts for nearly 10° of tibial external rotation as evidenced by the dial test [7] <i>(here, the authors missed to highlight the importance and clinical relevance of their key findings)</i>

tion; however, they fail to discuss the importance of those findings.

Fact Box 22.4: Conclusion

- The conclusion must be supported by the results.
- Make concise and accurate statements.
- Provide a sentence on the clinical or scientific merit of the study.

22.4 General Guidelines

An abstract is a 100–350-word snapshot of the manuscript, and it should not contain any information that is not further supported within the manuscript. While the abstract is the first, and in many cases, the only, portion of a research project that is examined by readers, it should be the last thing that is completed. Some may believe that an abstract should be written first since it is a short overview of the paper and immediately proceeds the manuscript, but it is much easier to summarize a manuscript that has already been completed. Additionally, it is more efficacious to write an abstract without concern for the word count and then pare it down to the specified word limit.

Within a limited space, it is easy for writers to misconstrue or bias the message of a research publication; thus, the writer must ensure that readers cannot misinterpret the abstract. Each component of the abstract should be able to stand alone such that the reader can clearly understand the message of each aspect without having to refer to other sections of the abstract for clarification. The abstract, and manuscript for that matter, should be written in the past tense and in the third person. For example, the sentence “The surgeon fixed the anterior cruciate ligament at the midpoint of the anteromedial and posterolateral bundle” should read “The anterior cruciate ligament was fixed at the midpoint of the anteromedial and posterolateral bundle.” Instead of phrases that incorporate “I” or “we,” phrases such as “the investigation demonstrates,” “the results illustrate,” or “this study explains” should be included in the manuscript. While it is appropriate to

include the aforementioned phrases in the abstract, they may be removed to save space for a sentence or phrase that helps convey the message of the abstract.

Abstracts do not include references, and it is also important to ensure that the abstract does not have any undefined abbreviations. Lastly, each abstract must contain a list of 4–6 “keywords” that will help guide the search indexes in finding this article.

Some authors recommend writing the abstract from scratch after finishing a manuscript, while some authors believe that it is best to copy direct phrases from manuscript to implement into the abstract. Those who write abstracts in the former mechanism believe that copying information directly from the manuscript leads to an abstract that is non-confluent or a summary that contains too much or too little information. In this method, it is best to reread the manuscript in its entirety and then summarize information in a new way that is unique from the original manuscript. Copying portions of the manuscript is an efficient method for creating the abstract since every piece of information lies within the manuscript itself. Ultimately, neither method is considered more efficacious than the other, but the modality in which the abstract is formed is dependent on the writer’s preference and their comfort level.

Fact Box 22.5: Abstracts for Scientific Articles

- Follow the instructions for authors
- Tell the same story as in the manuscript
- Focus on the data related to the purpose and hypothesis of the study
- The conclusion must be supported by the data and highlight the scientific merit

Take-Home Message

- An abstract is a snapshot of the manuscript and is a crucial part of each manuscript as it is freely available and frequently read.
- Abstracts should not contain any information that is not further supported within the manuscript.

- Regardless of the purpose of the abstract, it should contain the aim of the study, a brief overview of the methods, the most important results, and a conclusion addressing the clinical or scientific merit of the study.

References

1. Andrade C. How to write a good abstract for a scientific paper or conference presentation. *Indian J Psychiatry*. 2011;53:172–5.
2. Beaufils P, Becker R, Kopf S, Matthieu O, Pujol N. The knee meniscus: management of traumatic tears and degenerative lesions. *EFORT Open Rev*. 2017;2:195–203.
3. Canseco JA, Kojima K, Penvose AR, Ross JD, Obokata H, Gomoll AH, et al. Effect on ligament marker expression by direct-contact co-culture of mesenchymal stem cells and anterior cruciate ligament cells. *Tissue Eng Part A*. 2012;18:2549–58.
4. Cartwright R, Tikkinen KA, Vierhout ME, Koelbl H. How to write an ICS/IUGA conference abstract. *Int Urogynecol J*. 2010;21:509–13.
5. De Smet AA, Manaster BJ, Murphy WA, J. How to write a successful abstract. *Radiology*. 1994;190:571–2.
6. Forsythe B, Haro MS, Bogunovic L, Collins MJ, Arns TA, Trella KJ, et al. Biomechanical evaluation of posterior cruciate ligament reconstruction with quadriceps versus Achilles tendon bone block allograft. *Orthop J Sports Med*. 2016;4:2325967116660068.
7. Forsythe B, Saltzman BM, Cvetanovich GL, Collins MJ, Arns TA, Verma NN, et al. Dial test: unrecognized predictor of anterior cruciate ligament deficiency. *Arthroscopy*. 2017;33:1375–81.
8. Frazer A. How to write an effective conference abstract. *Emerg Nurse*. 2012;20:30–1.
9. Ickes MJ, Gambescia SF. Abstract art: how to write competitive conference and journal abstracts. *Health Promot Pract*. 2011;12:493–6.
10. Kopf S, Forsythe B, Wong AK, Tashman S, Anderst W, Irrgang JJ, et al. Nonanatomic tunnel position in traditional transtibial single-bundle anterior cruciate ligament reconstruction evaluated by three-dimensional computed tomography. *J Bone Joint Surg Am*. 2010;92:1427–31.
11. Li G, Abbade LPF, Nwosu I, Jin Y, Leenus A, Maaz M, et al. A scoping review of comparisons between abstracts and full reports in primary biomedical research. *BMC Med Res Methodol*. 2017;17:181.
12. Pierce LL. How to make a beef stew...or write a structured abstract. *Rehabil Nurs*. 2017;42:243–4.
13. Turbek SP, Chock TM, Donahue K, Havrilla CA, Oliverio AM, Polutchko SK, et al. Scientific writing made easy: a step-by-step guide to undergraduate writing in the biological sciences. *Bull Ecol Soc Am*. 2016;97:417–26.
14. Weinstein R. How to write an abstract and present it at the annual meeting. *J Clin Apher*. 1999;14:195–9.



How to Make a Good Poster Presentation

23

Baris Kocaoglu, Paulo Henrique Araujo,
and Carola Francisca van Eck

23.1 Introduction

Poster presentations are an important part of every scientific meeting [1, 17, 20]. Often new ideas and concepts are presented here [5]. A poster can be an excellent way to present a research project to an audience of interested peers and can be used to obtain feedback on a study [8, 16]. Peers can include fellow researchers but also surgeons, physical therapist, nurses, and engineers, and more [12, 19]. One major advantage of a poster presentation over a podium presentation is that a poster is available to be viewed during the entire duration of the meeting and can therefore gain more exposure [18]. Various types of poster presentations exist. Perhaps the most commonly known format is a printed poster displayed in an exhibit hall on a poster board (Fig. 23.1). However, more meetings are transitioning to electronic poster (e-posters). An e-poster is essentially a slide show presentation in which the slides advance automatically available on computers

distributed across the meeting (Fig. 23.2). Some meetings employ a combination of a physical and/or e-poster or an event to let the authors pitch the research presented on their poster with a short oral presentation. Regardless of the format, the poster should catch the attention of the audience while representing the study data in a clear and concise fashion [9].

Fact Box 23.1

A poster is an excellent way to present a research project and obtain feedback from peers.

This chapter aims to help orthopedic researchers in the preparation and presentation of a scientific poster. The learning objectives are to know the various different types of poster presentation, be familiar with the technical aspect of how to make a scientific poster, and understand what to do at the scientific meeting to get the most out of presenting research in poster format.

B. Kocaoglu
Department of Orthopedic Surgery, Acibadem
University Faculty of Medicine, Istanbul, Turkey

P. H. Araujo
Santa Luzia Hospital, Clínica COB, Brasília, Brazil

C. F. van Eck (✉)
Department of Orthopedic Surgery, University of
Pittsburgh Medical Center, Rooney Sports Complex,
Pittsburgh, PA, USA
e-mail: vaneckc@upmc.edu

23.2 Guidelines to Prepare a Poster Presentation

Because a poster is not designed as oral presentation, it should be prepared differently than a lecture. A poster should attract and engage the

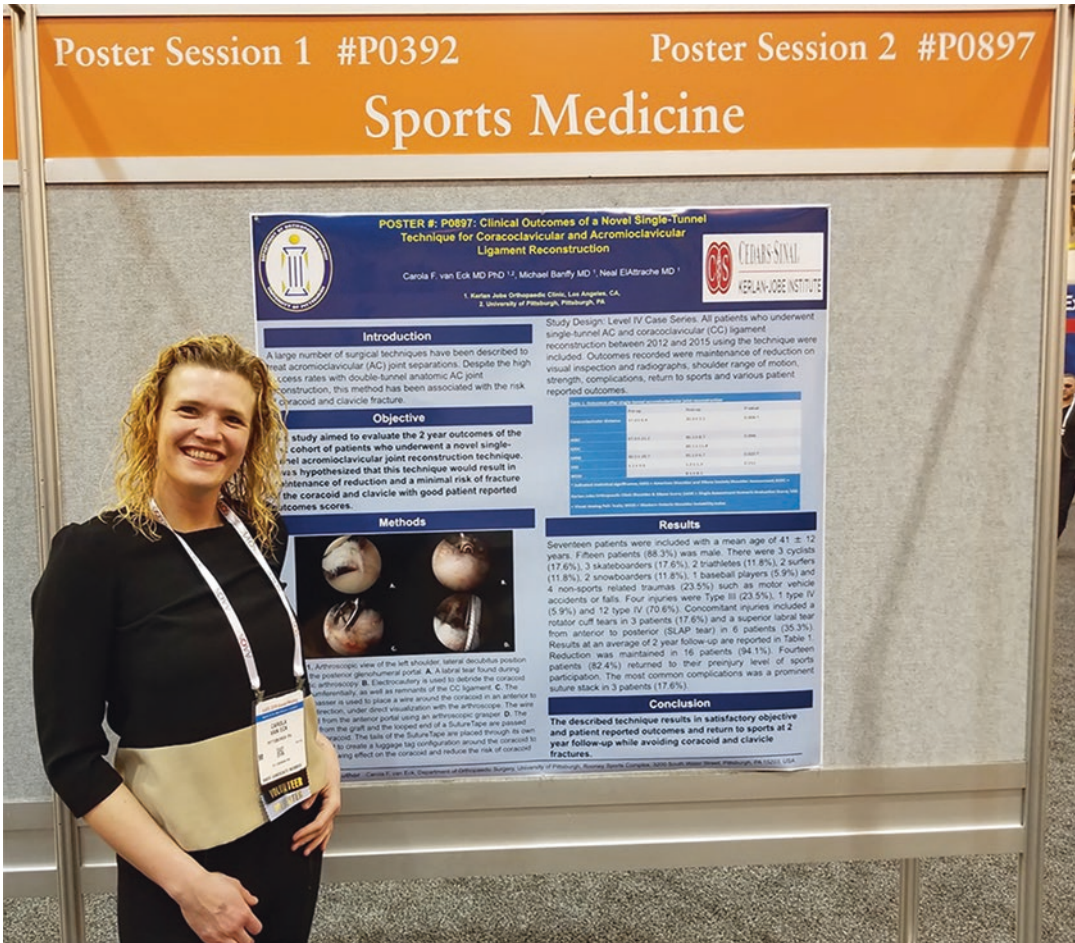


Fig. 23.1 Example of a conventional, printed poster displayed in an exhibit hall on a poster board

viewer by generating visual interest [2–4, 7, 10, 11, 13, 15, 21, 22]. However, when it comes to presentation of the data, this must be done on a stand-alone basis and be self-explanatory. This means that the readers of the poster should be able to understand the study aim, methods, results, conclusion, and significance, even when the presenting author is not there to explain anything. In addition, the figures must have clear figure legends and labeling where appropriate to facilitate this.

Generally, there should be a title portion, followed by objectives, methods, results, and discussion/conclusion section. Tables and/or figures can be used to allow for easier/more interesting presentation of the data. Conflicts of interest

should be disclosed, and contact information for the corresponding authors should be provided

Fact Box 23.2

A poster should be concise enough to attain the readers' attention and also complete enough to allow interpretation without verbal presentation.

[14]. Detailed instructions on how to prepare these sections are discussed below.

The title should be concise and attract the attention of people passing by. Oftentimes titles are too long. Phrasing the title as a strong state-

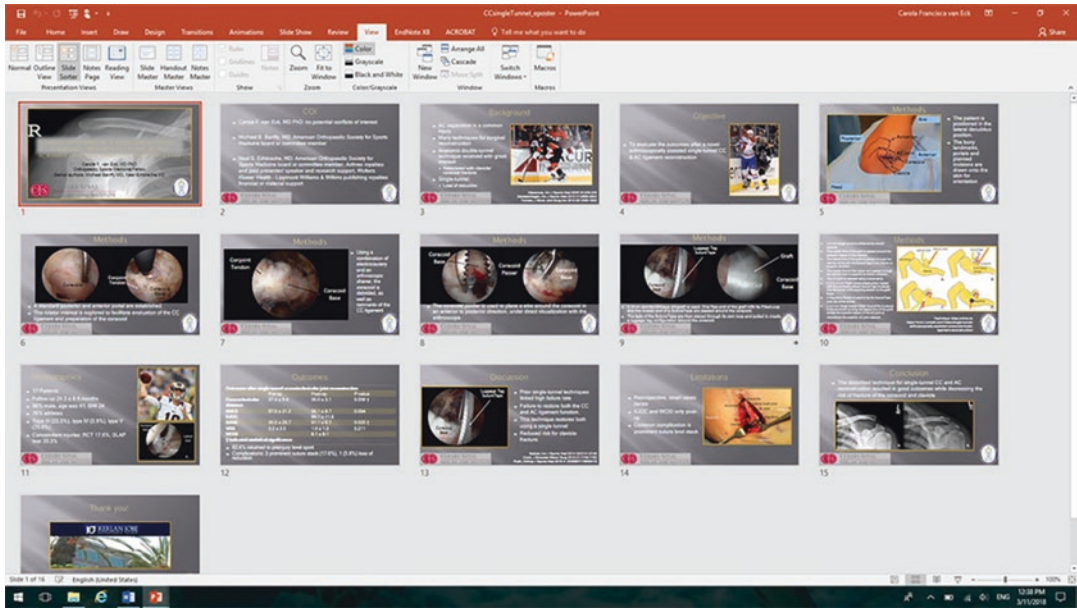


Fig. 23.2 Example of an e-poster. An e-poster is essentially a slide presentation in which the slides advance automatically on computers distributed across the meeting

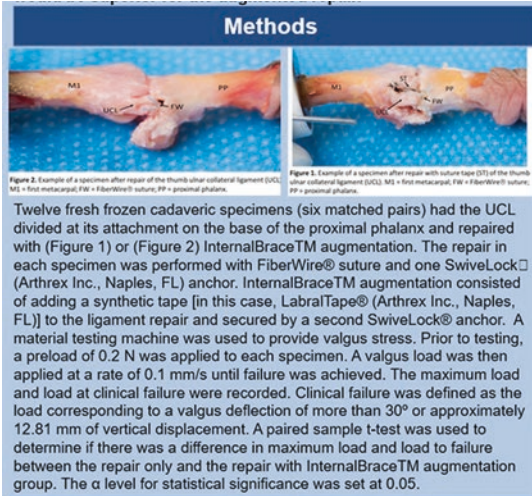
ment or a question is generally better to spark the interest of the readers. All authors with their credentials and the affiliated institution(s) should follow the title. If there are any conflicts of interest to disclose, this should also be done. The person in charge of making the poster should check the guidelines pertaining to how to disclose potential conflicts of interest from the specific meeting/organization.

The first text box should discuss the objective/hypothesis of the study. A short background may be provided if relevant for understanding the goal of the study. However, care must be taken to avoid making the poster too wordy and lose the attention of the reader. References are considered optional but again can be used sparingly if this is felt to be fundamental in understanding the rationale of the study. The methods section should be brief but present enough detail to understand the study design, nature, population, data collection, and statistical analysis. If tables or figures can be used instead of text, this should be strongly considered (Fig. 23.3).

Similar to the methods, the results are often better presented in table or figure format to catch and keep the reader's attention. Unlike in a manuscript, only the most important findings of the study should be described keeping this section short and concise (Fig. 23.4). The discussion/conclusion section should state the summary of the study. A brief discussion follows. The focus of the discussion should be on clinical relevance of the work presented, limitations, and future implications. Similar to the introduction section, the use of references is generally considered optional, and references should be used sparingly to decrease unnecessary wordiness of the poster distraction from the message of the research.

23.3 Technical Aspects

Perhaps equally important to the content of the poster is how well this content is presented [2–4, 7, 10, 11, 13, 15, 21, 22]. To start, the preparer of



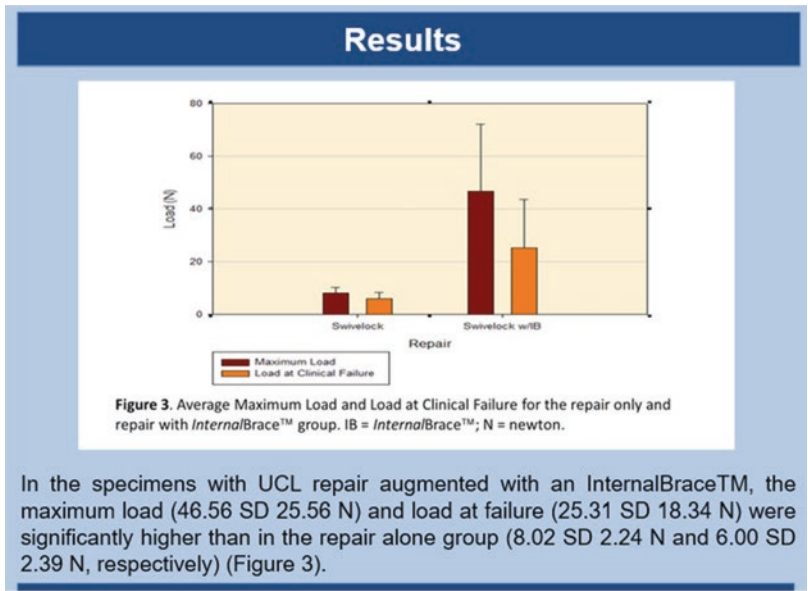
Methods

Twelve fresh frozen cadaveric specimens (six matched pairs) had the UCL divided at its attachment on the base of the proximal phalanx and repaired with InternalBrace™ augmentation. In the repair only group, the adductor aponeurosis was released and reflected volarly and the extensor mechanism dorsally. The joint capsule was incised at the dorsal border of the UCL. The UCL was detached from the volar base of the proximal phalanx and left attached at its origin on the first metacarpal. The guidewire for the suture anchor (3.5 mm DX SwiveLock® SL, Arthrex, Naples, FL, USA) was placed at the UCL attachment site on the volar base of the proximal phalanx. The guidewire was over drilled with a 3.2 mm drill. A 3-0 FiberWire® suture (Arthrex, Naples, FL, USA) was placed around the fork tip of the suture anchor and the anchor was placed into the drill hole, securing the suture in place. The suture was then passed through the distal UCL in horizontal mattress fashion to perform a direct repair of the ligament. In the repair with Suture Tape Augmentation Group the exposure was the same as for ligament repair only. The guidewire for the suture anchor was again placed at the UCL attachment site on the volar base of the proximal phalanx and over-drilled with a 3.2 mm drill. In addition to the 3-0 FiberWire® suture, a LabralTape® (Arthrex, Naples, FL, USA) was also placed around the fork tip of the suture anchor and the anchor was placed into the drill hole securing both the suture and tape in place. LabralTape® is a synthetic surgical tape that is currently being used to augment ligament repairs or reconstructions and provide internal bracing while biologic healing occurs.²² As was performed in the repair-only specimens, the FiberWire® suture was passed

Fig. 23.3 With regard to the methods section of the poster, use tables or figures instead of text where possible. The example on the left shows how figures are used to gain the readers’ interest as well as to reduce the text. In

the example on the right, the same information is conveyed using text only. Text is less appealing and harder to interpret

Fig. 23.4 Unlike in a manuscript, only the most important findings of the study should be discussed in the results section, keeping the result section short and concise. The example on the top lists only the key findings and uses an easy to interpret graph. The example on the bottom describes the same findings in text-only format. The latter is more difficult to interpret for the reader and may not generate as much visual interest



In the specimens with UCL repair augmented with an InternalBrace™, the maximum load (46.56 SD 25.56 N) and load at failure (25.31 SD 18.34 N) were significantly higher than in the repair alone group (8.02 SD 2.24 N and 6.00 SD 2.39 N, respectively) (Figure 3).

Results

The maximum load in the UCL repair with suture tape augmentation specimens (mean 46.56, standard deviation 25.56) was significantly higher than in the UCL repair-only specimens (mean 8.02, standard deviation 2.24) ($p=0.002$). The load at failure was significantly higher in the UCL repair with suture tape augmentation specimens (mean 25.31, standard deviation 18.24) than in the UCL repair-only specimens (mean 6.00, standard deviation 2.39) ($p=0.002$). Five of the six repairs (83.3%) incorporating the suture tape augmentation failed due to some degree of tape slippage, with one sample failing due to anchor pull-out. In the repair-only specimens, failure occurred due to a variety of different failure modes.

the poster should check the specific meeting to find out the dimension of the poster, which are allowed and recommended. Most posters are created using PowerPoint [Microsoft, Redmond, WA, USA] or something comparable. If you are part of a larger academic institution, hospital system, or research group, it may be worth checking if poster templates are in existence of your institution, which you can then utilize.

your institution. This can include the organization logo, picture, or slogan (Fig. 23.5). Conversely, this may benefit you as much as it benefits your institution, as the reputation of your institution alone may attract viewers to your poster. If no such template exists and you are the first person making it, try to pick a calm background color, perhaps matching your institution's logo colors, combined with a text color which stands out from the background [4]. For example, avoid yellow text on a white background. Refrain from using colors, which may be unconsciously perceived to be offensive such as red text. There is a fine line between attracting attention and the poster being a visual overload. The size and font of the text are also very important. It is best to use a neutral font that is easy to read from a distance, such as Arial or Sans Serif. Generally, the size of the letter is the largest for the title and section headings (at least >62), medium sized for the text (at least >44), and the smallest for the references and corresponding authors' contact information (>36) (Fig. 23.5) [2–4, 7, 10, 11, 13, 15, 21, 22].

Fact Box 23.3

Optimal visual poster presentation includes a calm background color and a neutral font which stands out from the background and is large enough to read from a distance.

The optimal format to present and promote your organization is to have a unanimous format that is easily recognized by others as belonging to



Fig. 23.5 The official template for ESSKA (European Society Sports Traumatology, Knee Surgery and Arthroscopy) congress. The optimal way to present and promote yourself, your poster, and your organization is to

have a unanimous format, which includes your logo. Use a calm background color and a neutral font with a color that stands out from the background

Lastly, if your poster is a printed poster, it will need to be printed ahead of the conference. Several companies are available online which perform scientific poster printing. You will need to upload your presentation file to their website, and generally you can select if you would like to see proofs before printing. This may be worthwhile when the poster contains pictures or graphs to ensure the resolution is high enough to provide a good-quality poster at the size that it needs printed on. Take into account that receiving the printed poster may take several weeks [3]. Allow extra time to reprint the poster if upon its arrival misprints or mistakes are identified and a new copy needs to be printed. Many large meetings now offer poster-printing services, which will have your poster ready for you at the meeting site. The major benefit is that this avoids the burden of traveling with several posters. However, the major limitation is that if you arrive to find there is something wrong with your poster, there will likely not be sufficient time for revisions. These pros and cons will have to be considered.

23.4 At the Conference

When arriving at the meeting, you will need to hang up (or upload) your poster. Be sure to check your poster numbers, location, and setup times on the specific conference website. Some meetings provide pushpins to hang up the poster, but this is not always the case, so it is best to check in advance.

Ensure that the presenter of the poster is available to stand with the poster at least during the mandatory time slots but more frequently than this if possible. Presenters can be disqualified for submission during future meeting if these rules are not obeyed. Popular times for people to view posters are during the (lunch) breaks and of course the poster sessions. Showing the poster to colleges or presenting it at a lab or research meeting at one's own institution prior to the scientific meeting may help generate comments and help the poster presenter be ready to answer any questions the audience might ask. Special judges may be assigned to score posters for awards. These are

usually based on the overall presentation, including the abstract, the poster itself, and the attentiveness and presence of the presenting author. The latter includes proper attire following the established dress code for the meeting. If unsure about the dress code, ask someone who has been to the meeting before or contact the meeting directly. It is always better to air on the side of being overdressed. Remember, you are representing yourself, your research, and your institution.

Fact Box 23.4

Proper dress code is important when presenting a poster as it reflects on yourself, your research, and your institution.

Meeting guidelines should be checked as poster presenters can be disqualified for submission during future meeting if rules are not followed.

A consideration can be to add a handout to your poster, which the readers may take with them. This could list the abstract, key points, and the contact information for the corresponding authors. Some meetings will have the abstract for each study available for registered attendees online, on a flash drive, or printed in the program book [6, 23].

Take-Home Message

- Poster presentations are a key component of any scientific conference.
- Often new ideas and concepts are presented here.
- Various types of poster presentations exist, including a printed format displayed in an exhibit hall, e-posters available on computers, and a combination of a poster with a short talk.
- A poster should attract and engage the viewer by generating visual interest.
- However, when it comes to presentation of the data, this must be done on a stand-alone basis and be self-explanatory.

- The person in charge of making the poster should check the guidelines for the meeting the poster is being presented.
- Although the content of the poster is important, so is the quality of the visual presentation.
- Choose previously used templates from your institution to ensure uniformity and easy identification of presentations related to the institution.
- Ensure that the presenter of the poster is available to stand with the poster during the mandatory time slots.
- Reviewing the poster with colleagues prior to the scientific meeting may help generate comments and improve the poster (presentation).
- The presenter should be dressed in proper attire.
- It is always better to air on the side of being overdressed.
- You are representing yourself, your research, and your institution [7].

References

1. Abicht BP, Donnenwerth MP, Borkosky SL, Plovanich EJ, Roukis TS. Publication rates of poster presentations at the American College of Foot and Ankle Surgeons annual scientific conference between 1999 and 2008. *J Foot Ankle Surg.* 2012;51:45–9.
2. Beal JA. Preparing for a poster session—some practical suggestions. *Mass Nurse.* 1986;56:5.
3. Boullata JI, Mancuso CE. A “how-to” guide in preparing abstracts and poster presentations. *Nutr Clin Pract.* 2007;22:641–6.
4. Briscoe MH. Preparing scientific illustrations: a guide to better posters, presentations, and publications. New York: Springer; 1996.
5. Daruwalla ZJ, Huq SS, Wong KL, Nee PY, Murphy DP. “Publish or perish”—presentations at annual national orthopaedic meetings and their correlation with subsequent publication. *J Orthop Surg Res.* 2015;10:58.
6. Donegan DJ, Kim TW, Lee GC. Publication rates of presentations at an annual meeting of the American Academy of Orthopaedic Surgeons. *Clin Orthop Relat Res.* 2010;468:1428–35.
7. Erren TC, Bourne PE. Ten simple rules for a good poster presentation. *PLoS Comput Biol.* 2007;3:e102.
8. Frank RM, Cvetanovich GL, Collins MJ, Arns TA, Black A, Verma NN, Cole BJ, Forsythe B. Publication rates of podium versus poster presentations at the Arthroscopy Association of North America Meetings 2008–2012. *Arthroscopy.* 2017;33:6–11.
9. Gundogan B, Koshy K, Kurar L, Whitehurst K. How to make an academic poster. *Ann Med Surg (Lond).* 2016;11:69–71.
10. Hamilton CW. At a glance: a stepwise approach to successful poster presentations. *Chest.* 2008;134:457–9.
11. Hand H. Reflections on preparing a poster for an RCN conference. *Nurse Res.* 2010;17:52–9.
12. Kleine-Konig MT, Schulte TL, Gosheger G, Rodl R, Schiedel FM. Publication rate of abstracts presented at European Paediatric Orthopaedic Society Annual Meetings, 2006 to 2008. *J Pediatr Orthop.* 2014;34:e33–8.
13. Lourie RJ. Preparing a poster presentation. *Nurse Educ.* 1989;14:10, 18, 23.
14. Matsen FA 3rd, Jette JL, Neradilek MB. Demographics of disclosure of conflicts of interest at the 2011 annual meeting of the American Academy of Orthopaedic Surgeons. *J Bone Joint Surg Am.* 2013;95:e29.
15. Miller JE. Preparing and presenting effective research posters. *Health Serv Res.* 2007;42:311–28.
16. Naziri Q, Mixa PJ, Murray DP, Grieco PW, Illical EM, Maheshwari AV, Khanuja HS. Adult reconstruction studies presented at AAOS and AAHKS 2011–2015 Annual Meetings. Is there a difference in future publication? *J Arthroplasty.* 2018;33(5):1594–7.
17. Ohtori S, Orita S, Eguchi Y, Aoki Y, Suzuki M, Kubota G, Inage K, Shiga Y, Abe K, Kinoshita H, Inoue M, Kanamoto H, Norimoto M, Umimura T, Furuya T, Masao K, Maki S, Akazawa T, Takahashi K. Oral presentations have a significantly higher publication rate, but not impact factors, than poster presentations at the International Society for Study of Lumbar Spine meeting: review of 1126 abstracts from 2010 to 2012 meetings. *Spine (Phila Pa 1976).* 2018;5:1347–54.
18. Preston CF, Bhandari M, Fulkerson E, Ginat D, Koval KJ, Egol KA. Podium versus poster publication rates at the Orthopaedic Trauma Association. *Clin Orthop Relat Res.* 2005;(437):260–4.
19. Schulte TL, Trost M, Osada N, Huck K, Lange T, Gosheger G, Holl S, Bullmann V. Publication rate of abstracts presented at the Annual Congress of the German Society of Orthopaedics and Trauma Surgery. *Arch Orthop Trauma Surg.* 2012;132:271–80.
20. Voleti PB, Donegan DJ, Baldwin KD, Lee GC. Level of evidence of presentations at American Academy of Orthopaedic Surgeons annual meetings. *J Bone Joint Surg Am.* 2012;94:e50.
21. White A, White L. Preparing a poster. *Acupunct Med.* 2003;21:23–7.
22. Wipke-Tevis DD, Williams DA. Preparing and presenting a research poster. *J Vasc Nurs.* 2002;20:138–42.
23. Zelle BA, Zlowodzki M, Bhandari M. Discrepancies between proceedings abstracts and posters at a scientific meeting. *Clin Orthop Relat Res.* 2005;(435):245–9.



How to Prepare a Paper Presentation?

24

Timothy Lording and Jacques Menetrey

24.1 Abstract Submission

The first step in presenting your research at a meeting is submitting an abstract. The call for abstracts often occurs quite early, especially for larger meetings, and can close a considerable time before the conference is to take place. For example, the call for abstracts for the 2019 ISAKOS congress closes in September 2018, a full 9 months before the meeting takes place.

Submission is usually online. Required information will include the author's details and affiliations, the title of your project, the abstract itself, as well as any financial disclosures.

The word limit for the abstract text can vary widely, from as few as 300 to as many as 800 words. Some meetings will dictate abstract sub-headings, but most are variations of the standard scientific structure: background, aims, methods, results, and conclusions. You may have already prepared a manuscript for your research, and if so

the abstract may be edited and repurposed for your submission. Use the word limit as much as possible, as it can be difficult to distil complex work into only a few hundred words. For each section of your abstract, determine the crucial points you want to convey. From this core information, expand your text to the word limit in a comprehensive manner. The submission process is competitive, and if English is not your first language, it may be worth asking a colleague to review your abstract prior to submission, to make sure your message is clear.

24.2 Presentation Structure and Preparation of Slides

Most free paper presentations are 6 min in length. Careful preparation is important, to ensure that the premise, findings, and relevance of your work are successfully conveyed in this short timeframe.

Preparing your talk and preparing your slides go hand in hand and for simplicity are considered here together. A paper examining the importance of tibial bony and meniscal slopes in anterior cruciate ligament injury is used as an example [1].

Similar to your abstract, most paper presentations will follow a standard structure, including an introduction, methods, results, discussion, and conclusion. Due to the time constraints inherent in a standard 6-min conference presentation, it is important to convey the most important

T. Lording (✉)
Melbourne Orthopaedic Group,
Windsor, VIC, Australia

The Alfred Hospital, Melbourne, VIC, Australia
e-mail: tlor ding@iinet.net.au

J. Menetrey
Centre de Médecine du Sport et de l'Exercice,
Hirslanden Clinique la Colline, Geneva, Switzerland
Service de Chirurgie Orthopédique et Traumatologie
de l'Appareil Moteur, University Hospital of Geneva,
Geneva, Switzerland

information clearly and concisely. Again, if you have already prepared a manuscript for your work, this will simplify the process of preparing your talk. Usually, we would estimate three slides per minute to be an appropriate pace for a scientific presentation.

Many institutions have slide templates available for presentations. If not, it is important to choose a template, colour scheme, and font that is easy for your audience to read. As a general rule, try to keep slides uncluttered, with a few main points per slide and clear, illustrative diagrams. Many presenters use slides with huge amounts of text or statistical information, often accompanied by an apology for using such a “busy slide”. The message is clearer if just the relevant information is presented. Try not to have more than eight lines of text on any slide and avoid small font sizes.

24.2.1 Title Slide

The title slide should include, at a minimum, the title of your work and the full authors’ names. Most would include the details of the conference, as well as the institutions from which the work arose (Fig. 24.1). Remember to acknowledge your co-authors when you introduce your work, and it is courteous to thank the conference organisation for the opportunity to present.

Most meetings mandate the second slide to be a disclosure slide, outlining any potential conflicts of interest in the work.

24.2.2 Introduction

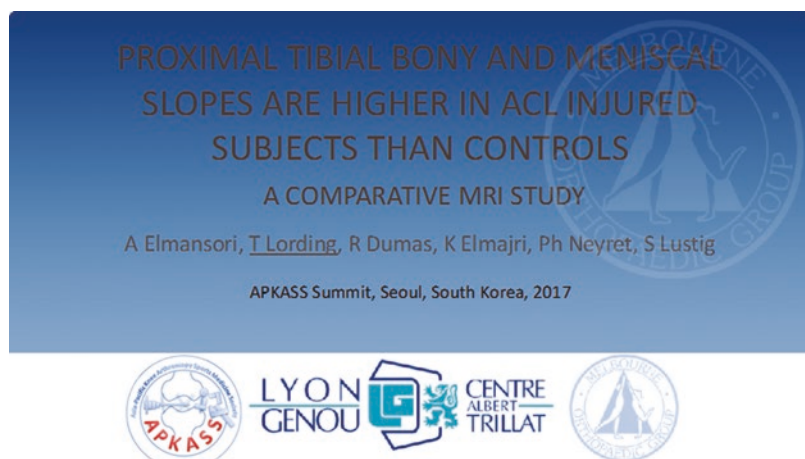
The introduction is critical in setting the context of your research. In two or three slides, or roughly 90 s, you need to convey to the audience the background to and purpose of your work, as well as your aim and hypothesis. More than any other section of your presentation, this should be tailored to your audience. For example, the audience at a general meeting may require more context to understand a paper about tibial slope in anterior cruciate ligament (ACL) reconstruction than would be required at an ACL-specific subspecialty meeting. Briefly, the current state of the science and the context of your research should be presented.

The scientific method involves generating a hypothesis and working to prove or disprove this hypothesis. It is important to clearly state the aim of your work as well as the hypothesis at the end of your introduction.

24.2.3 Methods

In this section, you should outline the materials and methods of your project. Due to the inherent

Fig. 24.1 Title slide



time constraints, this section must necessarily be a summary of rather than a fully detailed description of the methods used. At the same time, important points, such as demographic differences between treatment groups that might influence your results, should be highlighted and not glossed over. Photographs of testing rigs for biomechanical studies, or illustrations of radiological measurements, are particularly helpful in aiding the audience to understand your work (Fig. 24.2). Finally, any illustrations may help you to present the methods used in a succinct manner.

24.2.4 Results

The results section is often the shortest in a paper presentation. Frequently two or three slides is all that is required to present the main findings. Graphical representation is often helpful for your audience and more intuitive to understand than tables of numbers. Consider the data carefully when choosing the most appropriate illustrative style. For the tibial slope paper, a bar graph highlights the differences between groups well (Fig. 24.3). Don't present busy tables with dozens of numbers. Focus on the most important and pertinent findings of your results and present it in a way that leads on to your discussion. To add

clarity to your slides and to avoid detailed explanations, do not hesitate to add landmarks if you present anatomical or histological pictures. It is reasonable to offer to discuss your results further with interested parties after the presentation.

24.2.5 Discussion

The discussion section gives the greatest licence to the presenter to get the message of their work across. How does your work fit in with the findings of others? What do you think the main implications of your work are? Does your work help to explain the questions raised by other authors? Does your work raise new and interesting questions? These are the themes that may be explored in the time available and will vary from paper to paper. Briefly, your work should be compared to the current body of knowledge, using previous works, comparable investigations, followed by its originality, novelty, and clinical relevance. Your findings should be critically addressed point by point and hypothetical explanations explored. One trick is to write on post-it notes the different findings you may have to discuss and stick them on your computer or a board. It gives you a good overview and allows you to properly organise your discussion.

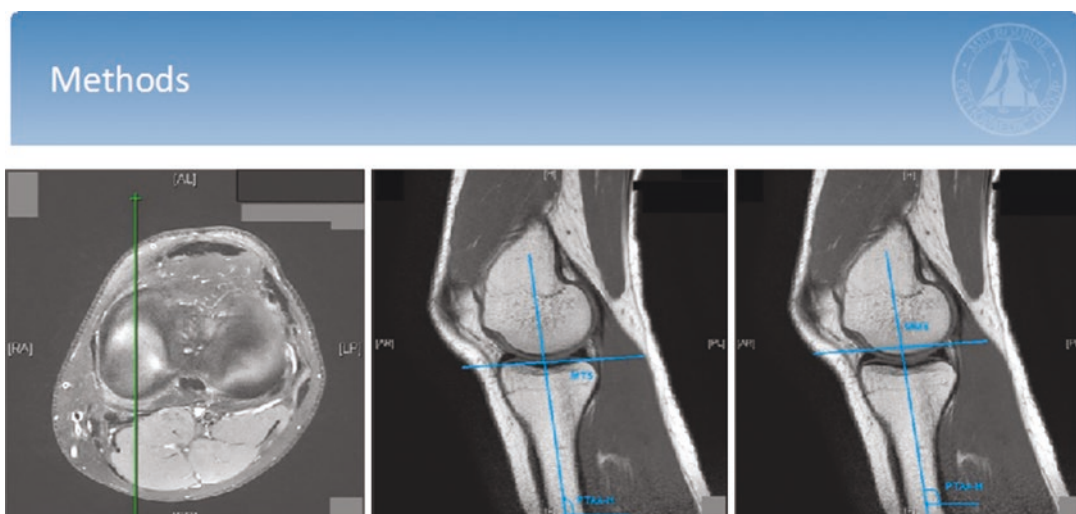


Fig. 24.2 Methods slide demonstrating measurement technique in the tibial slope study. Includes material from Elmansori et al. [1] (Reproduced with permission from Knee Surgery, Sports Traumatology and Arthroscopy)

Fig. 24.3 Results slide comparing lateral and medial bony and meniscal slopes. Includes material from Elmansori et al. [1] (Reproduced with permission from Knee Surgery, Sports Traumatology and Arthroscopy)

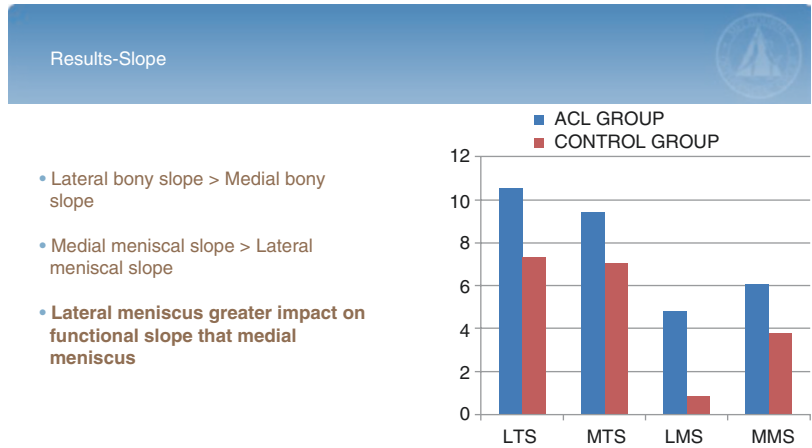


Fig. 24.4 Discussion of relevant results from the literature, highlighting the importance of the lateral tibial slope. Includes material from Stijak et al. [6] (Reproduced with permission from Knee Surgery, Sports Traumatology and Arthroscopy)

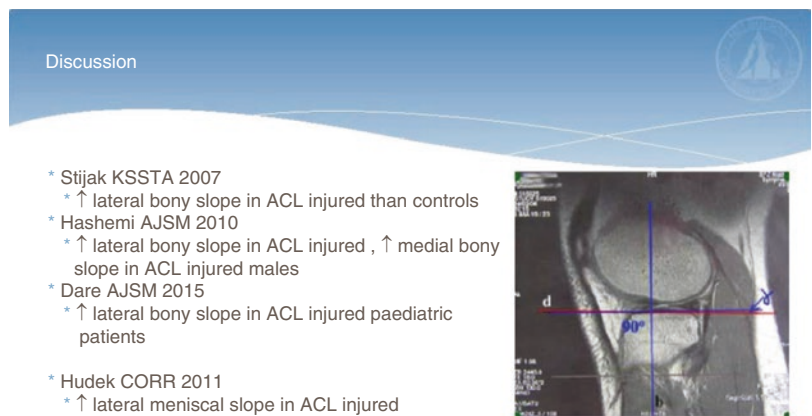
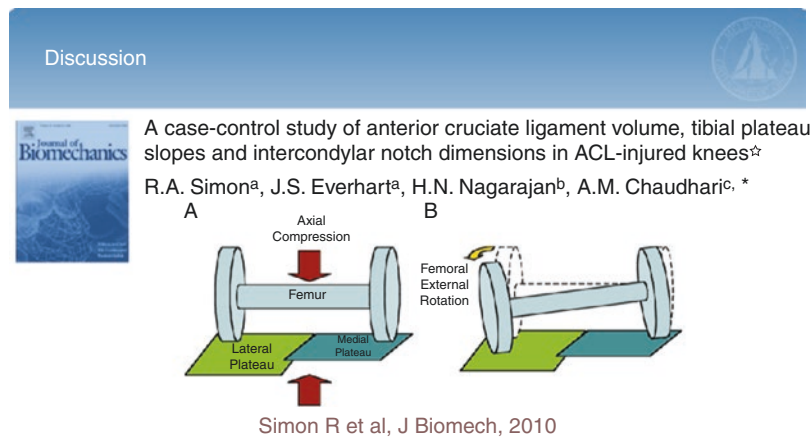


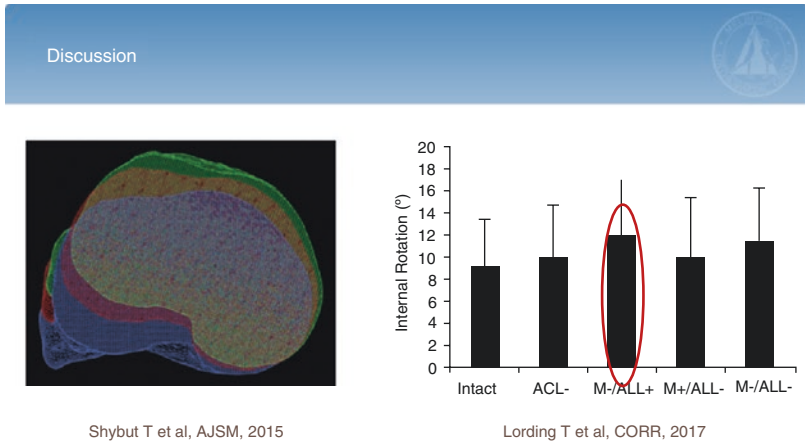
Fig. 24.5 Discussion of a postulated mechanism by which increased lateral tibial slope may increase ACL strain. Includes material from Simon et al. [4] (Reproduced with permission from the Journal of Biomechanics)



For example, in the tibial slope presentation, one slide was dedicated to the results of similar studies in the literature and highlighted the apparent importance of the lateral tibial slope, one slide discussed a postulated mechanism by which lateral compartment slope may cause ACL injury,

one slide examined the emerging evidence linking lateral meniscal injury to rotational instability, and one slide discussed the potential role for and current evidence for osteotomy to modify tibial slope in the ACL-injured knee (Figs. 24.4, 24.5, 24.6, and 24.7).

Fig. 24.6 Discussion of the role of the lateral meniscus in rotational instability. Includes material from Shybut et al. [3] (Reproduced with permission from the American Journal of Sports Medicine) and Lording et al. [2] (Reproduced with permission from Clinical Orthopaedics and Related Research)



Shybut T et al, AJSM, 2015

Lording T et al, CORR, 2017

Fig. 24.7 Discussion of the role of osteotomy. Includes material from Sonnery-Cottet et al. [5] (Reproduced with permission from the American Journal of Sports Medicine)

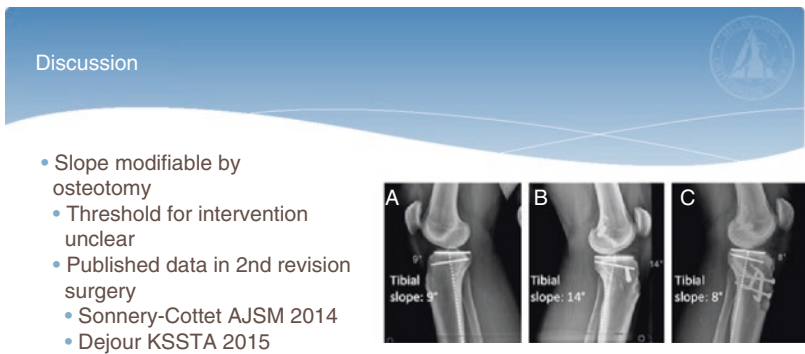
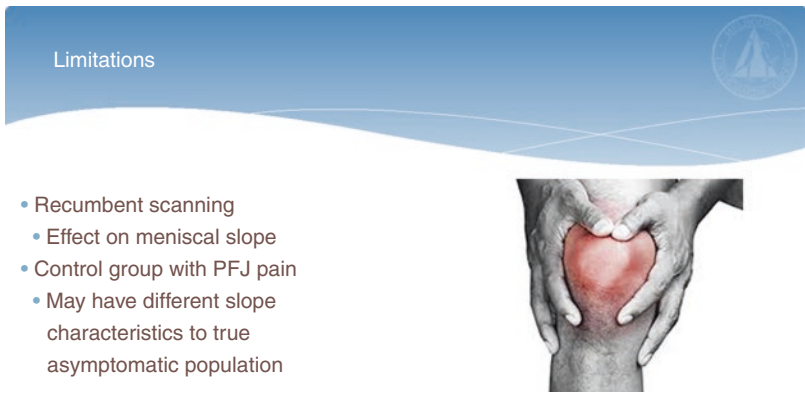


Fig. 24.8 Limitations slide



The discussion should conclude with a slide and brief mention of the limitations of your study and how these might influence your results (Fig. 24.8).

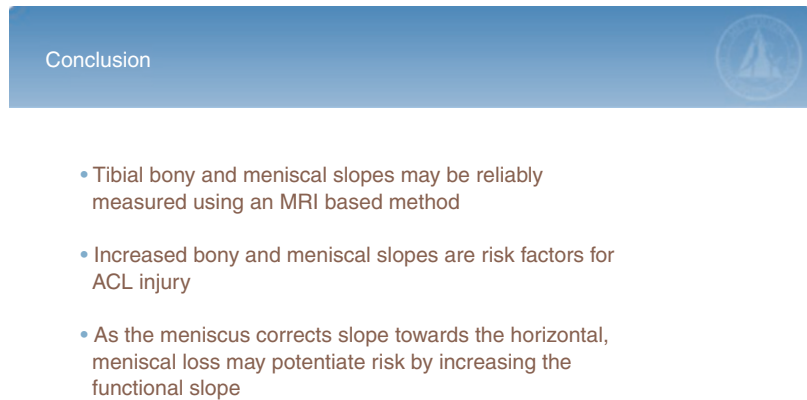
24.2.6 Conclusions

Clearly state the conclusions of your work in at most two slides. No new information or ideas

should be introduced at this stage, and ensure your conclusions are supported by your results (Fig. 24.9). Do not overstate findings you were expecting but you have not found. You have to find the right balance between advertising your results without overdoing it.

In a basic science presentation, a slide on which you report on future and ongoing works may prevent questions with respect to further development of your research.

Fig. 24.9 Conclusion slide with three clear points



Fact Box 24.1: Presentation Structure

Section	Slides	Content
Title slide	1	Title, authors, and affiliations
Declarations	1	Possible conflicts of interest Often prescribed format
Introduction	2–3	Background and purpose of study Statement of aims and hypothesis
Materials and methods	3+	Summary of investigative method Illustrations useful
Results	2–4	Short presentation of relevant result data Usually includes tables/graphs
Discussion	2–4	How does your work fit with previous work? What are the implications of the results? Does your work answer questions raised by others? Does your work raise new questions?
Conclusions	1–2	Clear statement of conclusions

24.3 Preparation for Questions

Taking questions about your research during the allocated discussion time can be the most difficult and stressful part of presenting your paper. As the questions may be unpredictable, this cannot be entirely rehearsed, but it is certainly possible to anticipate some questions and to be prepared.

Firstly, consider the work itself. As it is not possible to go into too much detail during the presentation itself, particularly regarding your methods, this is one area where the audience may ask for more details. Secondly, consider the work of others. Likely you will have discussed your work in the context of the results of other researchers. Make sure you are up to date with

relevant work, especially if there is a long lead time between abstract submission and your presentation. To this end, it is well worth while doing a literature search in the days leading up to the presentation, to identify any relevant new publications. Thirdly, consider not only the questions answered but also the questions raised by your work.

Lastly, take advantage of any available opportunities to practise your presentation. Many units encourage a “dry run” presentation within the group prior to larger meetings. This will allow you to practise not only the presentation but also to practise fielding questions and to see what type of questions your presentation stimulates. In two words: be prepared!

24.4 On the Day of the Presentation

24.4.1 Dress

It is important to dress appropriately for your presentation. In most circumstances, this would mean suit and tie for males and equivalent for females. Some meetings, such as smaller, sub-speciality meetings or meetings held at summer or winter resort locations, may have a more relaxed dress code. If in doubt, always err towards dressing conservatively.

24.4.2 Audio-Visual

You will need to find the speaker's room or audio-visual personnel to provide and upload your slides. In smaller meetings this may be as simple as going to the desk at the back of the room at the beginning of the session. Larger meetings often have a central speaker's room you need to locate. This is not always easy, especially at large meetings with many concurrent sessions, so make sure to arrive early to leave yourself ample time for this task.

Once your slides have been uploaded, review them to ensure that you have uploaded the correct version and that no formatting or layout errors have occurred. Always check that your videos are properly displayed on the organisational support and find out if they are automatically or manually played. Be prepared to comment on the content of a video if there are playback issues at the time of the presentation, which is not infrequent.

24.4.3 Delivery

Take a moment to stand at the podium before the start of your session to get your bearings. As a rule, on the podium you will have a view of your slides, as well as a clock or countdown for the time limit. Usually you will not be able to use a presenter's view showing the next slide and prompts, so if you feel written prompts will aid your presentation then prepare some small cards. Familiarise your-

self with how to advance slides. If you need a laser pointer, it is a good idea to bring one as one is not always provided. Adjust the microphone so it is close enough to pick up your voice but not so close that your speech is muffled. Inexperienced presenters tend to rush and speak quietly, so be conscious of speaking up, slowly and clearly. Similarly, when answering questions take a moment to consider before responding.

Take-Home Message

- Presenting a conference paper is an excellent way to share your research and ideas and to generate discussion with your colleagues.
- Presenting is intimidating, particularly for inexperienced authors and at large meetings, but experience is the best teacher and it gets easier with time.
- Preparation, rehearsal, and anticipation are essential to get your message across effectively and to portray your work in the best light.

References

1. Elmansori A, Lording T, Dumas R, Elmajri K, Neyret P, Lustig S. Proximal tibial bony and meniscal slopes are higher in ACL injured subjects than controls: a comparative MRI study. *Knee Surg Sports Traumatol Arthrosc.* 2017;25:1598–605.
2. Lording T, Corbo G, Bryant D, Burkhart TA, Getgood A. Rotational laxity control by the anterolateral ligament and the lateral meniscus is dependent on knee flexion angle: a cadaveric biomechanical study. *Clin Orthop Relat Res.* 2017;90:1922–8.
3. Shybut TB, Vega CE, Haddad J, Alexander JW, Gold JE, Noble PC, Lowe WR. Effect of lateral meniscal root tear on the stability of the anterior cruciate ligament-deficient knee. *Am J Sports Med.* 2015;43:905–11.
4. Simon RA, Everhart JS, Nagaraja HN, Chaudhari AM. A case-control study of anterior cruciate ligament volume, tibial plateau slopes and intercondylar notch dimensions in ACL-injured knees. *J Biomech.* 2010;43:1702–7.
5. Sonnery-Cottet B, Mogos S, Thaunat M, Archbold P, Fayard JM, Freychet B, Clechet J, Chambat P. Proximal tibial anterior closing wedge osteotomy in repeat revision of anterior cruciate ligament reconstruction. *Am J Sports Med.* 2014;42:1873–80.
6. Stijak L, Herzog RF, Schai P. Is there an influence of the tibial slope of the lateral condyle on the ACL lesion? *Knee Surg Sports Traumatol Arthrosc.* 2007;16:112–7.



Brendan Coleman

25.1 Why Publish?

There are many reasons why surgeons want to publish their research. For some, it is a requirement of their position, or they wish to enhance their career and gain promotion. For others, there is a moral obligation to the participants of the study to disseminate the knowledge gained and improve the knowledge of the orthopaedic community. There is an expectation for surgical trainees to undertake research. Publishing research can be helpful to learn the structure of scientific papers and enable surgeons to critique and identify the main points of other research. Although the prospect of writing a manuscript for publication can be daunting, the satisfaction of seeing your research published is worth the effort.

25.2 Before You Start

The best approach to research is to choose a topic that you are passionate about and that has relevance to your clinical practice. Writing a manuscript is hard, and it is the last 10% of effort that makes the difference in producing a publishable paper. If the topic does not interest you, it is easy to let the manuscript preparation slide when you

reach the final difficult stages. Don't be afraid of controversial topics as they can produce some of the most interesting papers [7]. Once you have chosen your topic, you need to formulate your research question and the hypothesis. The hypothesis is what you think the answer to your question will be. It does not matter if this is right or wrong, as the study is going to determine the answer.

Once a good research project has been identified, a thorough literature review should be undertaken. It is important to read the classic papers on the topic, but knowledge moves rapidly, and the relevant articles published over the last few years should also be reviewed. You need to determine if the research question has been investigated previously and what you are adding to the knowledge base on the topic.

It is helpful to seek the guidance of a colleague who has published scientific research in the early stages of planning your research. This will help in avoiding the common mistakes that are made in undertaking and publishing scientific research. Guidance could include identifying the right journal to submit your manuscript to. Once you have identified the journal, it is worthwhile reading the journal to understand the style of research and papers published [7].

B. Coleman (✉)
Department of Orthopaedic Surgery, Middlemore
Hospital, Auckland, New Zealand
e-mail: Brendan.Coleman@middlemore.co.nz

25.3 Types of Articles

There are a number of types of scientific articles that can be produced. These are discussed in further detail in Part V of this book. These include:

25.3.1 Case Report

This is the easiest place to start in producing a scientific paper with a short description of a rare or interesting case. It is uncommon for high-impact journals to publish case reports as there are few cases that have never been written about before. Clinical photos are often important in case reports.

25.3.2 Case Series

This is a review of a series of patients with the same condition who have been treated in a novel way. The data is often prospectively collected but retrospectively analysed. The STROBE initiative is an international collaboration of researchers and journal editors with the aim of improving the quality of research in observational studies [12]. By adhering to the STROBE checklist, you ensure that the reader can critically evaluate the validity of your study. The STROBE statement can be found at www.strobe-statement.org.

25.3.3 Case-Control Study

This is a type of study enabling comparison of two cohorts of patients who are matched in every aspect except the condition or intervention of interest. This can be a difficult study to perform as it requires careful planning to ensure the groups are well matched and that confounding factors are not introduced.

25.3.4 Controlled Study

This remains the gold standard of evidence-based medicine comparing an intervention group to a control group. The control group may be the

current gold standard treatment or no treatment. The participants are randomly assigned to either the intervention or control group. To ensure the study question is going to be answered, it is important that a power calculation is performed to ensure an adequate study size is undertaken. Controlled studies need ethical approval and registration and often need funding to perform.

25.3.5 Systematic Review or Meta-Analysis

These studies are an aggregation of high-quality randomized controlled trials published in the literature to investigate a specific question. There is a tendency in surgical systematic reviews or meta-analyses to conclude that the current evidence is not strong enough to make recommendations on treatment. The PRISMA guidelines are important when performing these types of studies (www.prisma-statement.org) [11].

25.3.6 Laboratory-Based Research

Basic science research is typically part of an ongoing research programme that aims to move into clinical research over time. These experimental studies can involve animals in research and usually need animal ethical approval.

Fact Box 25.1: Types of Articles

- Case report
- Case series
- Case-control study
- Controlled study
- Systematic review or meta-analysis
- Laboratory-based research

25.4 How to Write a Paper

Having completed the practicalities of the research study, it can feel like you have completed the research, but it is now that the critical element of the research takes place. Writing and completing the manuscript can be the most

challenging element of the study, and it can be easy not to complete the manuscript and publish the results.

The goal of the manuscript is to present the importance of your research and your findings. Writing needs to be precise and compact, utilizing short, simple sentences whilst avoiding complex scientific jargon [3]. You should be explicit and clear in describing the benefit of your paper.

The title of the paper should convey the impact of the results in a succinct manner. It should avoid reporting the results within the title. Randomized controlled trials, meta-analyses and systematic reviews require this to be in the title. “Five-year follow-up of unicompartmental knee replacement” is not very informative compared to “Increased risk of early revision in unicompartmental knee replacement”.

Writing a scientific manuscript is not like writing a novel. It requires a standardized manner of construction—introduction, methods, results and discussion (IMRAD) [3]. Although the majority of scientific manuscripts are written in English, many readers will not have English as their primary language [3]. It is imperative that the manuscript is easy to read and well organized and the language is simple and clear.

Fact Box 25.2

IMRAD method for writing manuscripts—introduction, methods, results and discussion

Writing the manuscript is hard and requires a rigid writing schedule to be put in place to ensure progress and completion. Follow the advice of George M. Whitesides, and start with a blank piece of paper, and write down, in any order, all the important ideas that occur to you concerning the paper [10]. Important elements may include the research topic and why is it important, the hypothesis, the major finding and other important findings.

When you create your first draft, let your ideas flow leaving any revision and editing for future drafts. As Paul Silvia advises, “your first drafts

should sound like they were hastily translated from Icelandic by a non-native speaker” [8]. Revising is a difficult skill, but the success of your paper is dependent on your ability to revise and edit. Experienced researchers make three times as many changes to the paper as novice writers [5]. One revision strategy is to learn your common errors and do a targeted search for them [2]. All writers have idiosyncrasies, and using the search function in your writing application may eliminate issues and improve the writing style [2]. Performing numerous drafts is typical when preparing a manuscript. Feedback from experienced researchers both within and outside your specialty field is crucial to improving your paper. Once feedback has been received, you can decide to adjust your paper as necessary.

25.5 Abstract

The abstract needs to be a stand-alone record of your research to entice potential readers into reading the full paper. With the prevalence of online publication databases, the abstract is more important than ever as a means of selling your paper. For many readers, the paper does not exist beyond the abstract [1]. Most journals require abstracts to conform to a formal structure and with a defined word count, typically 250 words. A structured abstract usually has sections on background, methods, results and conclusions.

The background section gives a very brief outline of why the subject is important and why the reader should care about the problems and the results. This section is typically two to three sentences.

The methods section should provide the reader with information to understand what was done. This should include study design, study population, treatment groups and a brief description of the treatment if needed and the primary outcome measure. The method section does not have to be in great detail.

The results section is the key element of the abstract, because readers of the abstract want to learn the findings of the study. The results section should contain as much detail as the word count

Table 25.1 Author checklist

<i>Introduction</i>
Are the objectives clear?
Is the importance of the study adequately emphasized?
Is the subject matter of the study new?
Is previous work on the subject adequately cited?
<i>Materials and methods</i>
Is the study population detailed adequately?
Are the methods described well enough to reproduce the experiment?
Is the study design clear?
Are statistical methods included?
Are ethical considerations provided?
<i>Results</i>
Can the reader assess the results based on the data provided?
Is the information straightforward and not confusing?
Are there adequate controls?
Are statistical methods appropriate?
<i>Discussion</i>
Have you commented adequately on all their results?
Have you explained why and how their study differs from others already published?
Have you discussed the potential problems and limitations with their study?
Are the conclusions supported by the results?

permits. It should include results of statistical analysis and secondary measures.

The conclusion section should describe the most important take-home message of the study. Other findings of importance can also be described. It is important that the conclusions match the findings of the paper and that the findings are not overstated.

Further information about writing of an abstract is contained in Chap. 22 (Table 25.1).

25.6 Introduction

The aim of the introduction is to present your research question and the importance of your research to the current literature. It should begin with a brief thought-provoking review of previous research on the subject, demonstrating why the topic is important, interesting or problematic in some way. It should create a niche for your study by indicating a gap in the current knowledge or how you are extending the knowledge on the subject. Once you have created the niche, the

author should outline the research question or hypothesis [9]. This needs to be explicit and clear in describing the hypothesis and where the potential benefit of the paper is to the reader. Although a relatively brief section, the introduction is vital to set the scene for the paper. Stay focused and concise—you do not want to answer questions of controversy in the introduction; leave that for the discussion.

25.7 Methods

The methods section of the paper should be the easiest part of the manuscript to write as it records what you have done during the study. The details of the study should have all been known prior to the commencement of the study. This section should include method of analysis and statistics but should not include the results. The methods section should be like a cookbook providing step-by-step instructions that can be repeated by others [7]. It requires enough accuracy and clarity to enable another researcher to repeat the study. Previously published research methods should be referenced if relevant [6].

Unless you have expertise in statistical analysis, advice should be sought from a statistician prior to commencing the study [3]. The most common error in research is to recruit too few patients leading to a beta error where no difference may be demonstrated [7]. It is critical to perform a power calculation prior to starting the study, ensuring you recruit adequate participants to demonstrate a difference in the treatment groups. Once the number of participants has been determined, an additional number of patients should be recruited to allow for loss to follow (transfer bias). The arbitrary minimum acceptable transfer bias determined by high-powered journals is 20% at 2 years [7].

Statistical methods are a necessary subheading within the methods section. You need to explain the statistical analysis used and information regarding the power calculation.

The study population needs to be clearly defined for the reader. It should be defined by providing clear inclusion and exclusion criteria.

The definition should include the time frame during which the study was undertaken. This should begin with the total number of patients eligible for the study followed by the number who met the inclusion/exclusion criteria. The number of patients followed up and the period of follow-up need to be described. Ideally, a maximum of 20% of patients should be lost to follow-up to make the data robust [7].

If there are diagnostic criteria for the inclusion in the study, these should be defined, e.g. patients were identified with rheumatoid arthritis, meeting the diagnostic criteria established by the American College of Rheumatology. Four of these seven criteria must be met for a patient to be classified as having rheumatoid arthritis: (1) morning stiffness, (2) arthritis of three or more joint areas, (3) arthritis of hand joints, (4) symmetric arthritis, (5) rheumatoid nodules, (6) serum rheumatoid factor and (7) radiographic changes.

Common problems with the method section include too long, too vague and not descriptive enough. Fluency of the text can be aided through consistency in the point of view in which it is written: first person “we” or passive voice. If writing in the first person, it can be easy to become repetitive in beginning sentences with “we did...” [13]. When revising the draft, ensure that there is variation in the beginnings of the sentences [13]. In writing the paper, the methods are often presented directly from the research protocol. This results in the methods being presented in the future tense yet the paper is being written on completion of the study and needs to be written in the past tense.

Approval from the appropriate institutional review board or ethical committee is an important check on the validity of the research design. In the methods section, ethical board approval should be documented.

25.8 Results

The results should be presented in a logical and orderly sequence. Initially, information about the study population and preoperative data should be presented. Operative findings and post-operative

outcomes follow, presenting early then late outcomes and complications.

In writing the results, the presentation of data should be mixed with text, tables and figures. “A picture is worth a thousand words” also applies to scientific papers, and a well-constructed figure or table is an excellent method to present multiple data points in a simple and effective manner. A table or figure should be a stand-alone representation of data. Although a succinct summary of the meaningful data presented in figures or tables should be included but only to highlight the important points rather than repeating the data already presented.

During research, often excessive amounts of data are collected. It is vital that the author is selective in the data that is presented. It must show the major findings of study but should discard any excessive data that does not add to the quality of the paper. Excessive data can be confusing and distracting for the reader, taking away from the important findings. Do not mistake this for falsification or manipulation of data which is unethical [6]. Rather you do not wish to present results that will not add to the current knowledge, guide clinical practice or are inconsequential. Avoid repetition in reporting results.

The perception of complications is different amongst surgeons, in part due to the impression that the treatment has failed but also due to the legal consequences in some jurisdictions. Good clinical practice means that complications or adverse events during a study should be reported. This enables information upon which shared decision-making can occur between patient and surgeon. Reporting of complications can be a vital element in altering techniques or implant choices. It is imperative that complications are reported in an unbiased manner.

In accordance with the good clinical practice guidelines of the International Conference on Harmonization, a serious adverse event is clearly defined as any untoward medical occurrence that:

- Results in death
- Is life-threatening
- Requires inpatient hospitalization or prolongation of existing hospitalization

- Results in persistent or significant disability/incapacity
- Necessitates medical or surgical intervention to prevent permanent impairment to a body structure or a body function
- Leads to foetal distress, foetal death or congenital abnormality or birth defect

Complications can be classified into treatment related or patient related. Treatment related can be due to surgical technique or a device or treatment. Patient related can be due to a local tissue condition or to the overall patient condition.

A statistically significant difference between treatment groups shows your study was powered sufficiently to determine a difference. However a statistical difference does not necessarily correlate to a clinical difference. For example, a statistically significant difference of three points on the Constant Shoulder score does not equate to a clinically relevant difference to the patients. Overlapping confidence intervals between groups indicate there is no clinically significant difference.

25.9 Discussion

The purpose of the discussion is to place the findings of your research into context with the wider knowledge of the subject. Due to the differing nature of studies, the discussion section will vary in its length and composition although the general structure remains similar. The structure of the discussion should start by recording the major findings of the study. The major findings should include the results of the hypothesis outlined in the introduction. The findings should be the message that you want the readers to take away from the paper—do not repeat every finding of the results section.

The second part presents a summary of the current literature. This mirrors the introduction but should not be repetitive. In the discussion section, the current literature expands on the major findings of the current study and helps place it within the wider body of knowledge on the topic. It is not a complete review of the lit-

erature but should compare and contrast your findings with those of other published studies, explaining any differences to previous research [4]. An explanation of the meaning and importance of the findings of your study should be presented. Consideration should also be given to alternative explanations for the results, particularly in relation to any unexpected or different findings. Although the importance of the major findings of the study may be clear to you, it may not be obvious to the reader [6]. Providing context of your findings in the wider knowledge of the subject gives direction to future research.

The third part of the discussion is to address the strengths and weaknesses of your study. It is especially important to address the weaknesses of your study as the reviewer and readers will be aware of the limitations. You should provide an honest critique of your paper's limitations, addressing the potential impact this has on your study's results and the relevance of the findings. Addressing potential issues with your paper will save you from comments from reviewers and leading to a negative impression of your paper by the reviewer or editor. It presents you as a thoughtful scientist. Where weaknesses exist, provide solutions or alternative explanations if possible. Discussing the limitations can create interest and pose further questions pointing to opportunities for future research.

In discussing the limitations of your study, it is important to recognize any bias within the study. Bias occurs in all scientific papers to some degree, and it is difficult to eliminate all bias. Selection bias occurs when the treatment groups are dissimilar. Prospective randomization minimizes selection bias as do strict inclusion and exclusion criteria. Performance bias occurs due to the person performing the study or treatment. An example is a single surgeon study introduces performance bias by the subtleties of their technique and experience that may not be reproducible to other surgeons. It is impossible to eliminate performance bias, as someone has to perform the research. Recording bias occurs if the data collector does not exhibit objectivity. This can be

Table 25.2 Common mistakes in writing manuscripts

Introduction and discussion too long
Lack of coherence and fluency in text
Overly long review of literature
Incorrect tense of methods
Lack of approval from institutional review board
Incomplete data
Not including detailed description of statistical methods
Poor quality figures, graphs and photos
Not discussing limitations
Concluding results beyond the study design

minimized with patient-reported outcomes or having the collector independent and blinded to the treatment received. Reporting bias considers how the outcomes of the study are reported. Using internationally recognized and validated outcome scores allowing comparison with other studies can minimize this [7].

The final section of the discussion involves a conclusion or the major take-home messages from your study. This should reiterate the answer to your research question and add the implications, practical application or recommendations from the study. The conclusion should only feature statistically significant findings although trends seen in the study can be reported in the discussion section (Table 25.2).

25.10 References

Journals can differ in the style that references are recorded, so read the author guidelines of the journal you plan to submit to prior to writing the reference section. The standard is 20–30 references for a scientific paper. The references should include the high-quality papers relevant to the topic and avoid low-quality papers if possible. Cite your own paper if required, but unnecessarily quoting your own papers in an attempt to improve citations detracts the quality of your paper. There are a number of electronic referencing tools available, and it is recommended that you utilize one of these tools during your manuscript writing. Examples of free bibliography managers online include Mendeley, Zotero and Citation Machine.

25.11 Manuscript Submission

Once you have completed your manuscript, put it aside for a week then reread, checking thoroughly tables and figures for accuracy and spellcheck. Then share it with your co-authors for comment and amend the paper as needed. Once you have chosen the journal, check the journal submission instructions then submit and wait. Very few papers are accepted immediately, and journals expect resubmission after corrections. Reviewers will comment on how the paper could be improved to allow publication. Revision of the paper taking into account the reviewers' comments and performed in a timely manner will enhance the chances of publication. If the paper is not suitable for publication, it will be made clear by the editor and reviewers. If the paper is rejected, then it should be revised taking advantage of the reviewers' feedback and then submitted to another journal. Achieving a publication is hard work, but it is rewarding to see your manuscript in print (Table 25.3).

25.12 Summary

Writing a manuscript for publication is a demanding task but, when successful, is a rewarding experience. The topic should be thought-provoking, sparking the interest of both the writer and reader. A winning paper starts with excellent preparation in determining the method and statistical analysis of the study. Writing should be

Table 25.3 Take-home points

Choose an interesting research topic
Structure your manuscript in the IMRAD method
Write in a clear, concise manner conveying the important information
Revise to eradicate errors and repetition
The abstract should stimulate the reader to read the paper
Explain what your research adds to the current knowledge
Discuss the limitations of your research to encourage further research
Persevere until your manuscript is accepted for publication

clear, focused and organized to allow the reader to gather the key points of your research that are relevant to the reader. The IMRAD style provides a format for the writer to follow. Ensure that discussion around the topic is relevant to the hypothesis and findings of your study. Be sure that the conclusions of your paper match the results of your study. Stay disciplined, work hard and you will be able to write a winning paper.

References

1. Andrade C. How to write a good abstract for a scientific paper or conference presentation. *Indian J Psychiatry*. 2011;53(2):172–5.
2. Belcher WL. *Writing your journal article in 12 weeks: a guide to academic publishing success*. Thousand Oaks: SAGE Publications; 2009.
3. Earnshaw JJ. How to write a clinical paper for publication. *Surgery*. 2012;30(9):437–41.
4. Faber J. Writing scientific manuscripts: most common mistakes. *Dental Press J Orthod*. 2017;22(5):113–7.
5. Faigley L, Witte SP. Analyzing revision. *Coll Compos Commun*. 1981;32(4):400–14.
6. Kallestinova ED. How to write your first research paper. *Yale J Biol Med*. 2011;84(3):181–90.
7. Karlson J, Marx RG, Nakamura N, Bhandari M, editors. *A practical guide to research: design, execution and publication*. *Art Ther*. 2011;27(Suppl 2):1–112.
8. Silvia PJ. *How to write a lot*. Washington, DC: American Psychological Association; 2007.
9. Swales JM, Feak CB. *Academic writing for graduate students*. 2nd ed. Ann Arbor: University of Michigan Press; 2004.
10. Whitesides GM. Writing a paper. *Adv Mater*. 2004;16(15):1375–7.
11. www.prisma-statement.org.
12. www.strobe-statement.org.
13. Zeiger M. *Essentials of writing biomedical research papers*. 2nd ed. San Francisco: McGraw-Hill Companies, Inc; 2000.



How to Write a Book Chapter

26

Thomas R. Pfeiffer and Daniel Guenther

26.1 Introduction

The invitation to contribute a chapter to a book can be perceived as an honor and appreciation of expertise. Most often, the invitation is based on prior scientific and clinical commitment. Writing a book chapter offers the opportunity to collaborate with colleagues, develop new or intensify existing connections, and spread expertise and ideas. A good book chapter fulfills the expectations and demands of the reader, is detailed but not lengthy, and fits into the context of the entire book. Especially at the beginning of a scientist's career, guidelines on how to fulfill the editor's expectations and how to catch the reader's interest are extremely helpful. This chapter aims to guide clinicians and researchers through the process of composing a book chapter successfully.

T. R. Pfeiffer (✉)

Department of Orthopaedic Surgery, Trauma Surgery and Sports Medicine, Cologne Merheim Medical Center, Witten/Herdecke University, Cologne, Germany
e-mail: t.pfeiffer@email.de

D. Guenther

Trauma Department, Hannover Medical School (MHH), Hannover, Germany
e-mail: guenther.daniel@mh-hannover.de

26.2 Prior to Writing or "Let's Get Started"

Organize yourself! Screen the letter of invitation for:

- Title and topic of the book and your assigned chapter
- Co-authors
- Timetable and deadlines
- Rough estimation of expected workload

Before you start writing, ask yourself: Do you feel comfortable with the given topic and is the expected workload manageable under given circumstances? Failing to meet a deadline and repeated requests of extensions beyond the deadline may damage your reputation, making future invitations and collaborations extremely unlikely. When the decision is made to accept the invitation, further information is shared by the editor and a more detailed screening of the author's guidelines should be executed [1]. This screening should address the following points (Table 26.1):

- Target readership
- Style of the book
- Outline of the book including all chapters

Books vary in style based on the target readership. Writing style should be adapted to their level of knowledge. Examples for different book

Table 26.1 Important information to consider prior to writing of a book chapter

Target readership	<ul style="list-style-type: none"> • Students • Residents • Fellows • Consultants • Researchers • Patients
Book style	<ul style="list-style-type: none"> • Primer/student teaching textbook • Specialist book • Surgical atlas
Book outline	<ul style="list-style-type: none"> • Chapter with subheadings to represent a group of themes (e.g., The Anterior Cruciate Ligament (ACL)) • Chapter about a special aspect of a topic (e.g., the double bundle technique of ACL reconstruction)

styles in the field of orthopedic surgery and sports medicine are, but not limited to:

1. *Primer/student teaching textbook*: A compendium of the most important facts and related clinical knowledge on the teaching topic is requested. The book should focus on the most important essentials. A short repetition of relevant anatomy and physiology is helpful.
2. *Specialist book*: Specialist books provide in-depth information on a certain topic. Based on most important current literature, state-of-the-art diagnostic tools or treatment methods are presented. A trend of increasing specialization can be observed over the last decade. While years ago, books with titles like *Sports Medicine* were published, nowadays more and more books with titles like *The Pediatric Anterior Cruciate Ligament* are published.
3. *Surgical atlas*: In the field of clinical orthopedics, surgical textbooks with step-by-step instructions are a popular format. High-quality images with virtual explanations, pitfalls, tips, tricks, and solutions from daily practice are requested.

It is helpful to know the book's outline and the other chapters of the book to prevent providing the same information twice.

As soon as you informed yourself about the frame conditions, contact your co-authors.

Identify the leading author of the chapter and the role of the co-authors. The leading author should guide the group through the writing process. Summarize the expectations of your co-authors with appropriate time lines. All authors should be on the same page and should be updated regularly through the writing process. At this point two approaches are feasible:

1. "The chain letter approach": One author writes a first draft of the chapter. This draft is forwarded to the next author, who adds content, revises the draft, and forwards the draft to another author. The order of authors can be adapted to experience and seniority. This reduces the number of turnarounds and saves valuable time.
2. "The melting pot approach": The work of writing is equally divided through the authors. The leading author should incorporate the different parts to the chapter and adapt the style to ensure the reading flow.

No matter which approach you choose, the final version should be double-checked and approved by all authors.

Talk about authorship early on in the process. Define who should be the first, second, or senior author. Sometimes, selecting the order can be uncomfortable. However, it will be even more uncomfortable if discussion about authorship starts after all the work is done.

Define deadlines for yourself and your co-authors. Missing deadlines will result in a rush prior to submission, which can ultimately lead to decreased quality of the chapter.

The author has to keep in mind that the book will compete on the market against comparable titles. Knowledge about what has already been published and what is lacking in current literature is inevitable. This knowledge can be achieved by reviewing published articles in journals and textbooks, scientific exhibits at conferences, and online versions. A good screening tool is PubMed (www.pubmed.gov) provided by the US National Library of Medicine and the National Institutes of Health.

Fact Box 26.1

1. Contact your co-authors! Identify the leading author of the chapter and the role of the co-authors.
2. Talk about authorship! Define who should be the first, second, or senior author.
3. Define deadlines for yourself and your co-authors.
4. Know what has already been published and what is lacking in current literature.

Fact Box 26.2

1. Come up with an outline of the book chapter before starting the writing process.
2. Outline can be elaborated as bullet point, catch phrases, or flowcharts.
3. Involve all authors to make sure that everybody is on the same page.
4. Use the framework of the composed outline to add text, piece by piece.

26.3 How to Fill the First Page(s) or “The First Step Is Always the Hardest”

The beginning of the writing process often poses a challenge and contains sticking points that young writers should be aware of. Frequently, one is tempted to write the first lines of a book chapter intuitively or based on personal strengths and interests. However, this approach involves the risk of losing the common thread. It is essential to come up with an outline of the book chapter before starting the writing process. Structuring should not be limited to subchapters and paragraphs, but the structure of each paragraph and the line of arguments should be drafted in the beginning. Based on personal preferences, the outline can be elaborated as bullet points, catch phrases, or flowcharts. It is helpful to have all authors involved in the process of tailoring the line of arguments in order to maximize the clinical experience and to include a broader spectrum of opinions. All authors will benefit from brainstorming, and everybody will be on the same page. This way, the structure of the book chapter will be maintained even if sections are written by different authors. If you do not approach your book chapter as a team, it is useful to involve other clinicians to make sure you do not get off track.

Once the outline of the book chapter is set, each aspect of the outline can be addressed by adding the corresponding information in the desired number of sentences.

26.4 The Writing Process or “How to Get the Job Done”

A good balance between theory and practice is key to an interesting and successful book chapter. Chapters consisting of mostly theoretical topics should contain implications for daily practice, tips, tricks, or annotations. In contrast to a descriptive review, a book chapter does not necessarily follow the regular outline of a journal’s article. Usually, there is no material and methods or results section [2]. Further, extensive statistical analysis to support the presented message is not needed. In general, at the end of a chapter, main findings, conclusions, and consequences for clinical practice should be presented.

The production of a book may take 2 or more years, and the reader wants a long-lasting “product”; therefore, only a selection of the most significant research performed in the past decade should be included. In contrast to a scientific journal article, expert opinions and clinical-based practices, as opposed to scientific or evidence-based practices, should be discussed. A book chapter does not require a full and complete representation of the existing literature [3].

Keep the terminology consistent to increase readability. In general, present tense and third person are used but can be altered in special needs. There is no need for alteration or use of synonyms for medical terms. A book chapter is still a scientific manuscript. Authors should avoid wordy sentences and relative clauses.

Usually, write the conclusion at the very end. It should include the most important take-home message of the article. Then, write the abstract. Afterward, double-check the title of the chapter. Does it still match the content of the chapter? Sometimes during the writing process, the central theme of the chapter changes stepwise, and at the very end, rewording of some initial parts is necessary.

Keep track of your references during writing. Sample reference programs are EndNote (Clarivate Analytics, Philadelphia, USA), Papers (ReadCube, Labtiva Inc., Cambridge, USA), or RefWorks (ProQuest LLC, Ann Arbor, USA). Travel libraries can enable the different authors to write on the same reference list.

Do not underestimate the impact of well-chosen figures. A good text is accentuated by thematically suitable figures. High-quality figures attract the reader's attention. In addition, they are a valuable tool to emphasize the take-home message. Especially for the purpose of presenting diagnostic and therapeutic algorithms, flowcharts and diagrams are useful to transfer theoretical knowledge to clinical work. Make sure that you provide figures with sufficient resolution and appropriate text reference. The related legend must explain the figure sufficiently. The reader should not have to return to the main text to find explanations. All abbreviations used in the figure must be defined in the legend.

Clear and consistent naming is needed to keep track of drafts. The naming should include year, month, day, running title, name, and version (e.g., 2017-11-11-ISAKOSResearchBook-PfeifferV1.docx). Each new draft should be saved. This enables authors to backtrack the development of the manuscript. In addition, make sure to save your work regularly during the writing process to prevent losing data.

Dropbox (Dropbox Inc., San Francisco, USA) can be used to share your drafts. Another option is to work in a cloud. Google Cloud (Google LLC, Mountain View, USA) or iCloud (Apple Inc., Cupertino, USA) provide options to work online on one version with multiple authors. However, keep in mind that you as the corre-

sponding author are responsible for keeping track of all changes and staying organized.

Read the final version of the book chapters multiple times before finalizing it. If the book chapter is not in your first language, a revision by a native speaker is highly recommended to ensure a professional and scientific writing.

Fact Box 26.3

1. Chapters consisting of mostly theoretical topics should contain implications for daily practice, tips, tricks, or annotations.
2. A selection of the most significant research performed in the past decade should be included.
3. Keep the terminology consistent.
4. At the very end, rewording of some initial parts is necessary.
5. Keep track of your references during writing.
6. High-quality figures attract the reader's attention.

26.5 Technical Considerations

Important technical points to consider are:

- Word count
- Number and requirements of tables, figures, and references
- If applicable, online version/videos
- Author instructions

Most book chapters contain approximately 10–15 manuscript pages. A manuscript page contains about 4500 characters. Editor and publisher plan total number of pages and assign different number of pages to the authors. This needs to be strictly adhered to. Double spaced, font size 12, and Times New Roman or Arial are regularly requested styles. Do not spend time on excessive formatting as this will be performed by the publisher [4].

Tables and figures are usually saved as separate files. The technical requirements of figures differ distinctly between different publishers. Generally, a resolution greater than 300 dpi for figures is requested. As aforementioned, high-quality figures enhance the quality of a manuscript.

Before final submission, all technical requirements should be double-checked. When the editor requests revisions and makes suggestions to improve the draft, all changes should be highlighted. Proofreading is the final step of the writing process. The manuscript should be checked for misspellings and proper grammar. Additionally, particular attention should be turned to authors' names and institutions.

Take-Home Message

- Writing a chapter is a team effort.
 - However, the lead author, who must be defined prior to writing, needs to organize the group, keep everybody on the same page, and ensure the flow of the chapter.
 - Knowledge about what has already been published and what is lacking in current literature is imperative.
 - It is important to come up with an outline of the book chapter before starting the writing process.
 - Then, the framework of the composed outline can be used to add text, piece by piece.
- Keep in mind that a book should be a long-lasting product and competes on the market against comparable products.
 - In general, a reader is attracted by scientifically written text with consistent terminology that is supported by well-chosen, meaningful figures.
 - A coherent train of thoughts is key to a successful book chapter.
 - Chapters consisting of mostly theoretical topics should contain implications for daily practice, tips, tricks, or annotations.
 - Finally, a printed version of the book will reward authors for the sometimes intense, but always interesting, work of chapter preparation.

References

1. Lewerich B, Götze D. What to do when you are asked to write a chapter. In: Troidl H, McKneally MF, Mulder DS, Wechsler AS, McPeck B, Spitzer WO, editors. *Surgical research*. New York: Springer; 1998.
2. Kendirci M. How to write a medical book chapter? *Turk J Urol*. 2013;39(Suppl 1):37–40. <https://doi.org/10.5152/tud.2013.052>.
3. Woodrow L. Publishing research: book chapters and books. In: *Writing about quantitative research in applied linguistics*. London: Palgrave Macmillan; 2014.
4. Skipper T. Writing an effective book chapter. A guide for authors working with the National Resource Center for The First-Year Experience & Students in Transition; 2011.



How to Write a Winning Clinical Research Proposal?

27

Christian Lattermann and Janey D. Whalen

27.1 Introduction

In the modern clinical research environment, it has become a necessity to communicate efficiently, clearly and with a certain amount of enthusiasm in order not to lose the audience, be that patients, fellow researchers or a captive student audience. It is imperative that a research proposal, grant application or scientific paper tell a captivating story, containing the traditional elements of research rationale, hypothesis, study design, material, methods, etc. Hence, the ability to package a cut and dry research proposal into a captivating proposal is arguably the single most important and equally so underemphasized skill during scientific training in the medical profession.

This chapter will attempt to explain and illustrate how this can be achieved by dividing the chapter into a more narrative general section and a rather dry, to the point second section.

While this is a somewhat “different than usual” approach, it serves a distinct purpose. Both parts are necessary to create a scientifically relevant, valid and interesting proposal.

C. Lattermann (✉)
Brigham and Women’s Hospital, Harvard Medical School, Boston, MA, USA
e-mail: clattermann@bwh.harvard.edu

J. D. Whalen
Icartilage.com, Boston, MA, USA

27.2 Section 1

27.2.1 Funding Mechanism

- *The introduction to any scientific research proposal should be targeted to the actual funding mechanism.*

Funding mechanism generally provides a clear description of what the funding entity is looking for. This information is often contained in the request for application (RFA) but sometimes must be sought out specifically on specific websites, locations or mission statements. While the language used in these funding announcements can often be used more efficiently to put your infant to sleep, it is important to assure that your research proposal carries a laser focus on the description of the funding mechanism. Elements such as target population, mechanism, disease, burden to society or sometimes even materials and methods may be very specifically defined inside the request for proposals (RFP). Often the request for application will use specific language, phrases or terms that offer themselves to be repeated in your proposal. Identifying these phrases and using them frequently will assure that there is no doubt about the relevance of your research proposal to the identified funding mechanism. (e.g. the grant mechanism is interested in funding proposals that aid the “translation” of scientific data into the

clinical realm; the key word is “translation”. This word should be identified and clearly emphasized in the research description.)

- *The science is meritorious; the story, however, needs to be flexible.*

The scientific question that you are asking may not be completely aligned with the question that is being asked by the RFP. However, most research proposals ask more than one scientific question, i.e. there is more than one specific aim. If an RFP asks for “injury prevention techniques in female football players” and your research proposal is targeting all football players, it is easy to include a specific aim that specifically identifies factors for female soccer players rather than the entire population. Data collection may still be carried out on the entire population, but the proposal will seek funding for just a subgroup of your overall scientific aim and is focused on the RFP. This flexibility may help you to broaden your relevance and impact.

- *Read the RFP carefully and with attention to detail.*

Every RFP is written differently and asks for different elements, page lengths, addenda and prerequisites. Some require ethics committee proposals to be submitted prior to submission; some require investigators to be tied to a university; some require “citizenship” (in the USA); some require certain membership status (foundation grants); and some do not. It is important to read these descriptions carefully prior to embarking with a proposal. Writing a successful grant proposal is hard work, and the last thing you want is to find out in the 11th hour that you are not eligible.

- *Look at the deadlines and be realistic, and create a master timetable.*

Generally, smaller grant funding mechanisms will ask for short, brief proposals, 1–2 pages,

sometime less, and sometimes they can be a quad chart only. While the actual writing process for those kinds of proposals may look easy, you may find out quickly that *a shorter proposal is not necessarily easier to write*. In fact, the old saying by Blaise Pascal, “if I had more time, I would have written a shorter letter”, holds true also for grant proposals [1].

It is hard to suggest an ideal amount of time for a proposal, as this always depends on how much previous work you have done and how much pre-existing language and sections you already may have in your portfolio. If you are writing your first proposal in your specialty, give yourself a minimum of 2–3 months for a small grant funding mechanism. Larger foundation grants and NIH-style grant mechanisms or cooperative grant mechanisms asking for 6 or 7\$ figure grant support over multiple years require sometimes several years for a team of writers to prepare a competitive proposal. A good rule of thumb is to identify the shortest amount of time you think it will take you to write a complete first draft and double that number. It should also be kept in mind that a good and winning proposal requires review by yourself as well as by peers. This process may take 4–6 weeks. One strategy may be to push as hard as possible towards a deadline, intentionally miss the deadline in the last minute and then spend the next funding period fine-tuning and polishing the proposal for a submission to the next RFP cycle. There is no shame in not submitting a proposal to a first deadline. Resubmission into the next cycle is usually possible, and a well-written proposal that has been reviewed by peers and polished is a win-win for everybody even if it gets rejected. It can serve as a starting point for a new proposal or may be adapted into a paper or presentation. An unfinished proposal is a waste of time, as it will not likely be funded or serve future proposals.

To track your timelines, it is advisable to create a master timetable with soft deadlines for certain sections.

Fact Box 27.1

A good rule of thumb is to identify the shortest amount of time you think it will take you to write a complete first draft and double that number. It should also be kept in mind that a good and winning proposal requires review by yourself as well as by peers. An unfinished proposal is a waste of time, as it will not likely be funded or serve future proposals.

27.2.1.1 How to Approach the Actual Preparation of the Proposal

- *Learn to work in sections. No proposal gets written in a day.*

Grant sections can generally be divided into the key elements containing the rationale, specific aims, statements about innovation and relevance and study design. Another essential and often more voluminous part is comprised by the “technical” pages.

While it is important to have a clear research question, hypothesis and specific aims, it is often the technical pages that will cause the biggest problems when time is short. These technical pages (i.e. standard operating procedures, inclusion of women and children statements, etc.) are short, usually one or two paragraph statements, for which you or your institution may already have pre-existing language. This language may just need to be adapted to a specific proposal. These “technical” pages do not require ingenuity but nevertheless need to be prepared. This is work that can also be delegated, or, if not, then this is the type of work that may be done in-between surgical cases, before or after a clinic or lectures or on an airplane or in an airport lounge.

- *Have a clear view of the science, relevance, innovation and study design before you start writing the specific aims page.*

It is advisable to work first on the hypothesis and the specific aims and develop the scientific

rationale from there. Then expand into the innovation and relevance section. Once those three sections are written, bits and pieces of those sections can be combined and rearranged to write the introductory section and an abbreviated rationale of the specific aims page. Hence, the specific aims page is usually finished after the other sections have been written and requires the largest time commitment.

- *Do your prep work.*

Several aspects of a proposal are addressed under what one can consider “prep work”. After you have gathered your thoughts, you should try to put those together in a brief chart or text. This can then serve as the famous “elevator pitch”. The scientific equivalent to this is the “napkin sketch”, a brilliant way to scrutinize your idea and discuss it with peers and team members over a lab meeting. During my basic science training in the Ferguson lab in Pittsburgh under Chris Evans, we had an unwritten rule. If you could sketch out your project on a napkin during the Friday afternoon lab meeting and convince your peers of the value of the project, you would get the go ahead to pursue this project. While this is a light-hearted version of the basic idea, it illustrates a fundamental concept. A brief and to the point illustration with few words requires careful thought and a clear understanding of the subject matter. This exercise helps to focus any research idea. Some grant proposals require this type of sketch in the form of a “quad chart”.

Based upon the “napkin sketch” or rough outline of the project, an in-depth actual research protocol should be developed. While it is important not to forget about the other parts of the grant proposal, this protocol is critical, as it may influence several other sections such as the biostatistics section and feasibility of the project and may identify numerous hurdles that need to be addressed and that may subsequently alter the actual proposal itself.

Another helpful concept is to create a separate file for conceptual lines of reasoning, explanations, supportive literature, etc. for the main sections.

These can be addressing significance, relevance, economics, innovation, future directions, etc. Every investigator constantly reads literature about the specific topic of interest, and even though access is almost unlimited, many times great ideas to justify a hypothesis suddenly show up and disappear the next day. Creating a small folder on the phone or the computer allows for a quick repository of these ideas that can be recalled when needed.

Involvement of a trained biostatistician from the beginning is a critical step for all clinical proposals. Even for animal-based basic science research, it has become mandatory in many institutions. The estimated sample size, power calculations and potential study design considerations should be made early on to design the proposal correctly. It is easier to do this upfront rather than at the end, as the study design affects the actual specific aims, time lines as well as budgeting.

Research your internal submission deadlines and include those on your master timetable. Most institutions require you to run any grant submission to be run through a research office. Each office has different time requirements to approve a grant project. Sometimes there may even be internal institutional decisions about eligibility for submission that may exclude an investigator from a grant submission upfront. Hence, it is important to inform your institution about a grant submission early on and start the internal process early.

27.3 Section 2

The second section of this guideline will address the subsections of a proposal and how to package scientific information in a captivating and interesting fashion.

27.3.1 Significance/Relevance of Your Research

The significance and relevance statements come with substantial time requirements. It is not advisable to use well-known and worn-out lines

of argumentation. Amongst reviewers, there is a “significant and relevant” (!) amount of fatigue that sets in when proposals use the same few phrases repeatedly to justify their relevance (i.e. “Over 200,000 ACL reconstruction are being performed in the United States annually”).

It can become quite comical from the perspective of a reviewer when most of proposals have the same sentence in the relevance statement and then even cite the same paper. The one proposal that will introduce a different approach becomes instantly attractive. Hence, the significance/relevance statement often benefits from the author spending some time researching lesser known but relevant information obtained from epidemiology papers on the topic. This information can often be found in lesser known, non-orthopaedic sources such as the Centers for Disease Control and Prevention (CDC), governmental white papers or scientific articles published by authors from a different profession, who may have a different perspective of the problem. Try to address the actual “knowledge gap” and previous failures to solve the problem and/or new aspects that highlight the importance of your topic or line of research.

27.3.2 Innovation

This may be one of the most important sections of the grant proposal, as many grants are seeking to fund “highly innovative” proposals. How does one portrait a proposal as “highly innovative”?

While there is no gold standard of innovation, it is worth pointing out several commonly seen arguments towards innovation of a project that are usually not appreciated by a critical reviewer. I will list those, briefly and without judgement:

- Incremental changes on an existing measurement technique or technology to measure an already described and published event or pathology
- Being in possession of a tool that is looking for an application
- Material testing of a new device or commercially available proprietary technique
- Retrospective data analysis

An innovative proposal instead should answer the following questions:

- Why are you or your team the group to do this research?
- What barriers can you break down that have not been broken down yet?
- Are you communicating, reaching across “silos” or have a unique team?
- Are there aspects of your proposal that can only be done in your environment?
- Which scientific connections can you make in this proposal that are truly unique?
- Be careful with statements such as “we show for the first time” but rather point out the scarcity of data and approaches to the problem that you are tackling.
- Outcomes tools/instruments should be reported with their descriptions (i.e. literature referencing their purpose, validation, MCID, MDC).
- Study time points and group determinations.
- Study endpoints (a study endpoint is the most important outcome criterion as success and failure to answer your hypothesis are determined by the study endpoint).
- Anticipated difficulties, failure mitigation strategies and limitations of the study.

27.3.3 Material/Methods/Study Design

This is the most technical part of your proposal and should be written as such. Hence, there is no need to “tell a story” but stay short, brief and to the point. In the following is a short bullet point list of elements that should be include in your material/method section:

- Pre-studies to determine sample size, enrolment numbers and feasibility.
- Brief mention of the labs’ or clinics’ capacity (the main part of this information will be under resources, but it is helpful to put it into perspective; e.g. in an ACL study, it is helpful to mention that patients are being recruited from a clinic that performs over 300 ACL surgeries a year).
- Subject flow and experience (i.e. how are patients being recruited? Is this part of an ongoing registry or cohort recruitment effort?).
- Strategies to prevent drop-outs and patients that are lost to follow-up.
- Strategies to prevent bias.
- For some grant mechanisms, it is important to create prospective enrolment charts to assure diversity. This can be included in the M/M section.

27.3.4 Biostatistical Support Is Important

As mentioned above, it is important to obtain input from a biostatistician early in your study planning as it can affect technical aspects of your study. Specifically, it is important to address the following issues in your statistical plan and data management plan:

- Power analysis
- Randomization
- Bias
- Calculation/estimation of potential study “drop-outs” and estimation of patients “lost to follow-up”
- Estimate or verify enrolment numbers (may require a pre-study to determine those numbers)
- Missing data management plan
- How to deal with outliers
- Specific interim analysis time points
- Statistical analysis techniques that may influence the study design format for clinical trials (i.e. adaptive designs, umbrella trial platforms, etc.)

27.3.5 References

Nowadays reference managing software is mandatory for a scientific paper or proposal.

It is important to comply with the rules about references. Our group once submitted an NIH proposal with 16 references instead of 15 (as was called for in the request for proposals), and the proposal was rejected.

A few bullet points about this topic are as follows:

1. Use a reference managing software.
2. Be current (i.e. last 3–5 years unless citing a landmark paper that is relevant).
3. Do not inundate the reviewer in unnecessary references.

27.3.6 Not Funded and What Now?

Finally, a word about how to deal with the decision of your proposal. There are two forms of rejection:

1. *Technical rejection for failure to comply with submission guidelines, etc.*
2. *Failure to convince the reviewers of the value of your work*

Make sure you keep things in perspective. Remind yourself that the odds of getting funded are relatively slim. Funding lines may be as low as single digit percentiles, and you are competing against the best researchers in your field.

The technical rejection is usually consequence of not having read the instructions properly. You will get a response quickly and typically prior to the actual review process. This hurts but is precisely the reason why we point out in the chapter how important it is to read the instructions. Depending on how competitive the mechanism is, these technical rejections may happen easily.

Failure to get funded should be rather the norm and is not to be considered a failure of the project or the author personally. Given the fact that most of these funding cycles deal with hundreds of proposals written by smart and well-trained individuals, it is unrealistic to think that you will have a proposal accepted the first time around. While this may happen, more commonly, a failure to get funded results in another set of excellent critiques.

Having gone through this process many times myself, I would recommend accepting the rejection letter, read it and, put it aside. These letters always hurt and create frustration. It is common to think that the reviewer just did not understand

the topic or your proposal, that they did not like you or your team and that the critique is not fair. A seasoned researcher will step away for a few days and then come back to the letter and dissect the critique carefully. It is important to understand that this critique has been written for two reasons. First, it is to point out scientific weaknesses in the proposal. These are usually fixable problems and give you an opportunity to substantially improve your proposal. Second, the reviewers are trying to guide you towards what they expected and would like to see return to the study section or grant review committee. This insight is extremely valuable and allows the researcher to adjust a future proposal according to those recommendations. In certain cases, the proposal was identified as simply misguided for the mechanism in which case this is good advice, as a resubmission will not likely be successful.

Fact Box 27.2

There are two forms of rejection—technical and failure of relevance. Make sure you keep things in perspective. Remind yourself that the odds of getting funded are relatively slim. Accept the rejection letter, read it, and put it aside. A seasoned researcher will step away for a few days and then come back to the letter and dissect the critique carefully.

Take-Home Message

- Write to a lay audience trained in the scientific method but not in your specific scientific discipline.
- Be prepared and take your time.
- A successful proposal should often be revised prior to success.
- Use all resources available and discuss your proposal with colleagues.

Reference

1. Shapiro FR. The Yale book of quotations, section: Blaise Pascal. New Haven: Yale University Press; 2006. p. 583.



How to Review a Clinical Research Paper?

28

Neel K. Patel, Marco Yeung, Kanto Nagai,
and Volker Musahl

28.1 Introduction

The ability to critically review a paper is a valuable skill to have, not only for the peer-review process of a journal but also to be able to interpret findings of a paper in the context of clinical practice. However, typically there is no formal training on how to review a paper throughout medical school or residency. Thus, this chapter will serve to provide general guidelines to use when reviewing a paper.

Each section of a paper provides important information regarding the purpose, study design, findings, or interpretation of the results. As a reviewer, having a set of guidelines or questions in mind while reviewing each section can be useful to ensure that nothing is missed in the process of the review. The questions to ask and the items to evaluate can vary depending on the study design or article type. This chapter will review what should be present in each section of a paper and what questions should be asked within each section. The main sections of a paper to be discussed are the introduction, which provides the aims of the study, the methods, the results, and the discussion, which gives an interpretation of the results. The guidelines outlined in this chapter will serve as a good foundation that reviewers

can adapt to specific papers and individualize based on their particular preferences.

28.2 Assessment of Research Aims and Purpose

Generally, the introduction states the issues related to the topic of the paper and formulates the rationale for the questions and hypotheses of the study. The organization of the introduction may differ depending on whether the paper is a clinical report, a study of new scientific data, or a description of a new method. Most studies are published in order to (1) report entirely novel findings, (2) confirm previously reported work (i.e., case reports, small preliminary series) when such confirmation remains questionable, or (3) introduce or address controversies in the literature when data and/or conclusions conflict [1]. One of these three purposes generally should be apparent in this section. Usually, the first paragraph introduces the general topic and/or problem and suggests its importance. The second (and third) paragraph provides the rationale for each question or hypothesis, and a final paragraph states the questions and hypotheses. The rationale should be established by providing representative literature and places the rationale of the work in the context of the current body of

N. K. Patel · M. Yeung · K. Nagai · V. Musahl (✉)
Department of Orthopaedic Surgery, University of
Pittsburgh, Pittsburgh, PA, USA
e-mail: musahlv@upmc.edu

literature. The reviewer should critically check the following points:

- Does the introduction provide the necessary background; in other words, why is it necessary to do this study?
- What is known/unknown about the topic?
- Is the research question of interest? Is it original or merely a repetition of existing knowledge? Does the manuscript itself have a new message?
- Is the purpose clear? Is the hypothesis given? Is there any clinical relevance in a basic science study?
- Can the introduction be shortened? More often than not the introduction can be shortened without losing the main message.

If you are not familiar with the topic of the paper, you should search the previous literature. This will provide you with information about what is known/unknown and what is controversial about the topic of the paper. As reviewer it is important to assure that pertinent and up-to-date literature is cited while at the same time respecting some of the classic literature. Additionally, it is important here to assure that the current study is truly novel and not a duplication of a previously published study.

28.3 Reviewing the Methods of a Clinical Research Paper

The methods section is perhaps one of the most important sections to assess when reviewing a clinical research paper. Poor methodology can lead to suspect or flawed results and conclusions. The fundamental principles of critically appraising this section are ensuring transparency, clarity, and reproducibility of the methods and maintaining internal validity by minimizing any sources of bias in the study.

Authors should describe in detail and in a clear and logical sequence how the study was designed, executed, and analyzed. A reader should be able to accurately reproduce the study based on the author's description within the

study's methods. The design of the study should be clearly described and justified: Was the trial carried out prospectively or retrospectively? Was the design a randomized trial, a case control design, or a case series? Furthermore, any obtained institutional review board or ethics committee approvals should be declared if applicable. Similarly, any study registries such as clinical trial registrations or systematic review registration should be stated.

The eligibility criteria of the patients or subjects in the study should be clearly delineated. All inclusion criteria such as patient demographic factors, diagnoses, etc. should be specified. Inclusion criteria based on diagnoses should be defined by specific symptoms, objective clinical findings, or radiographic parameter. In a study including patients with femoroacetabular impingement, the author's definition for a positive diagnosis should be stated; for example, patients were diagnosed with cam-type impingement in the presence of hip/groin pain, positive flexion, adduction, internal rotation impingement testing, and positive radiographic parameters such as alpha angle greater than 50°. Similarly, all relevant exclusion criteria such as age, comorbidities, and concurrent diagnoses should be determined a priori and clearly stated.

A study's interventions must also be well described to allow reproducibility. In the case of a surgical study, a detailed, step-by-step description of the surgical technique similar to a comprehensive dictated operative report should be provided. Furthermore, prior conservative management, surgical indications and decision-making, postoperative rehabilitation, and restrictions should be reported. Similarly, studies with medical interventions must include reporting of all dosages, frequency of therapy, and duration of therapy.

The outcome measures used in the study should be described by the authors, with clear delineation of the primary outcomes that the study seeks to assess versus secondary outcomes. Authors must not only explain what the outcomes are but how they are assessed, as well as the timing at which they are assessed. If clinical outcome measure or scores are used, these measures should

be validated for the corresponding diagnoses and patient population. For example, in early hip arthroscopy research, some patient-reported outcome (PRO) scores initially designed and validated for hip arthroplasty were used to assess surgical outcomes in younger, hip arthroscopy patients. With development of novel patient-reported outcome scores such as HAGOS (Hip and Groin Outcome Score) or IHOT (International Hip Outcome Tool), these newer measures validated for use in young to middle-aged adults with hip and groin pain may be more appropriate for use in the femoroacetabular impingement and hip arthroscopy demographic [4, 7, 8].

28.3.1 Statistical Analyses

The assessment of appropriate statistical analyses can often be overwhelming. The reporting of descriptive statistics is important to consider when reviewing a paper. Typically in studies that have a large sample size without outliers, the mean and standard deviation should be reported. On the other hand, if the study sample size is small and there is outlier with the sample, a report of the median and range of results is preferred to provide the reader with a better understanding of the sample. Calculations to determine sample size should be included in relevant clinical trials to ensure the study is adequately powered. The authors should disclose the statistical software that was used for analyses. The type of tests used for statistical analyses for each comparison should be discussed, and the reviewer should determine if these are appropriate for the type of variables analyzed in the study. Analyses of continuous variables should be performed using the appropriate tests such as student's *t*-test, whereas analyses of parametric variables should be performed using tests such as the X^2 test or Fisher's exact test. Studies comparing the means of more than two groups should use analysis of variance (ANOVA) testing if comparing one independent variable and two-way ANOVA or multivariate ANOVA testing for multiple independent variables. If multiple secondary comparisons are made, appropriate Bonferroni

corrections should be performed to adjust for the multiple comparisons.

28.3.2 Study Design Specific Methodology

Core principles in assessing study methodology have been discussed above; however, there are different salient points to assess depending on the type and design of study being critically appraised. For example, when analyzing a randomized controlled trial, it is important to scrutinize allocation concealment and blinding, whereas in a retrospective cohort study, this may not be relevant. The following section will discuss important methodological considerations specific to different study designs.

28.3.3 Randomized Controlled Trials

The key feature of randomized controlled trials is that their purpose is to control bias through randomization and blinding. The difference between an effective randomized controlled trial and a poorly executed one can be attributed to its methodology in implementing these features. The execution of the randomization process should be clearly defined. The specific method used to generate the randomization sequence should be discussed, as well as the type of randomization and relevant methodological details such as blocking.

The method of allocation concealment is an important factor to consider when appraising the methods of a randomized controlled trial. Many strategies of allocation concealment can be fraught with potential for selection bias. Sealed envelopes with randomization data within it can be subverted by investigators by holding it to the light to look through it or even by discretely opening the envelopes. Using seemingly unbiased determinants like randomizing chart numbers can cause bias if investigators, knowing their treatment allocation, choose to include or exclude patients based on their perceived outcomes. The literature cites many ways investigators can decipher the randomization results if

inadequate allocation concealment strategies are used [5]. Centralized randomization is the gold standard method of allocation concealment. In this method, the central office is called to establish study eligibility, at which time the randomization allocation for the included patient is given to the investigator.

Blinding is an important process to protect randomization after the assignments to interventions has been made and is another consideration in assessing a randomized controlled trial. Blinding of subjects to their assigned intervention can remove any bias of preconceived notions or subject behavior that could affect the results of the study. Blinding can be extended to the care providers, outcome assessors, and other study personnel to further reduce bias. The paper should explain which parties were blinded following assignment to intervention and how this was managed and maintained. However, one must consider that sometimes it is near impossible to blind certain parties such as blinding a surgeon to the intervention they are performing—a characteristic that is unique to surgical studies. The CONSORT (Consolidated Standards of Reporting Trials) guidelines and checklist are an invaluable resource to direct critical appraisal of randomized controlled trials (<http://www.consort-statement.org/consort-2010>) [6].

28.3.4 Systematic Reviews

The search strategy is an important key to reproducibility of the systematic review. The strategy can be quite complex in order for the authors to thoroughly, yet efficiently identify all relevant studies to include in the systematic review. The authors' search strategy, including all search terms and search language, should be outlined in a way that it can be repeated. Furthermore, all information sources (databases, etc.) used should be described, including the date that the search was last performed. With regard to study selection, clear inclusion and exclusion should be delineated. Preferred Reporting Items for Systematic Reviews

and Meta-Analyses (PRISMA) guidelines also recommend a flow diagram to summarize the inclusion and exclusion process, including the number of studies included and excluded at each step of the review and the reasons for such. The review process to screen studies for inclusion can include a title and abstract review, followed by a mandatory full-text screening. This process should be done in duplicate (by two independent reviewers) in order to reduce errors of excluding studies that in fact meet the study criteria. Intra-reviewer agreement statistics should be divulged to the reader to disclose how often consensus or disagreements occurred, and subsequently, methods of resolving any conflict surrounding inclusion/exclusion should be reported. Any algorithms or attempts to identify potentially overlapping study data should be discussed.

Data extraction preferably should be performed in duplicate to minimize error. Any attempts to obtain data from original authors not published in the included studies should be reported. All data points or variables that authors sought to extract from the included studies should be stated, even if they were not available or present in all studies.

Risk of bias and study quality of the included studies should be assessed by the authors. This can be performed using various tools, scales, and checklists. A commonly used tool is the Cochrane Risk of Bias tool, which assesses various domains of possible bias in the included studies, including selection bias, performance bias, detection bias, attrition bias, reporting bias, and other biases [2].

If a meta-analysis is performed, the choice of the summary measure (e.g., relative risk) should be stated and justified. The statistical method should be discussed, including the decision of use of random effects models versus fixed effects model, with authors providing rationale for their choices. Heterogeneity or consistency of the results between studies should be analyzed.

The PRISMA statement and checklist are an invaluable resource to direct critical appraisal of systematic reviews (<http://www.prisma-statement.org>) [3].

28.3.5 Biomechanical Studies

The biomechanical testing method must be described in detail, in a fashion that allows reproducibility of the study. Any jigs or testing apparatus should be described, with diagrams where appropriate to clearly communicate the experimental setup. Furthermore, any parameters of biomechanical testing should be outlined clearly, including any external loading conditions applied, as well as the quantifiable amount and direction of force applied. The use of cadaveric models requires reporting of specific age, type of preparation, and medical or surgical history of the included specimen. Similarly, the use of any robotic tools or apparatus used for evaluation of kinematic forces should be clearly described. All outcome parameters of biomechanical testing should be delineated, whether it be bending, torsional, or tensile strength of an orthopedic implant or kinematic forces measured across a joint.

Fact Box 28.1

Assessment tools that can help while evaluating the methods of a paper:

- Randomized Controlled Trials
 - CONSORT guidelines
- Systematic Reviews
 - PRISMA guidelines/checklist
 - AMSTAR checklist

28.4 Scrutinizing the Results of a Paper

The results should mirror the methods as described, including the statistical methods. The points that should be checked are as below:

- Is the results section organized in the same order as the methods section?
- The numbers including variables should be reported.

- All outcome measures mentioned in the methods should be reported.
- The text should be consistent with tables and figures. Text can usually be shortened or deleted in favor of tables and figures.
- Depending on the type of study, the results section may contain the demographic data instead of the methods section. In that case, check whether this data is clearly provided or not.
- Are the results adequately reported, with numbers (not only percentages) and distribution values, like standard deviation (SD), standard error (SE), or confidence interval (CI)? Is the reproducibility of the measurements (i.e., intra- and/or inter-observer intraclass correlation coefficients (ICC)) reported?
- Are the appropriate number of significant figures used based on the repeatability of the measures?

28.4.1 Statistical Significance

Statistical significance is only a guide and it cannot address whether findings are important for clinical outcomes. In clinical medicine, confidence intervals are often more helpful than statistical significance. For example, a study investigating the effect of the treatments A and B on the length of hospital stay may conclude that there is a statistically significant difference ($P = 0.001$) in the length of stay between treatments A and B. However, if the length of stay is 4.1 days with treatment A and 4.7 days with treatment B, it is not clinically meaningful even though it is statistically significant. This example indicates that statistical significance is not necessarily equal to clinical significance.

On the other hand, when the result shows “no significant difference” between the groups, the analysis should also be assessed because there is a possibility of type II error (β). It could be due to insufficient power of the test ($1 - \beta$), and the question of whether a power analysis (sample size calculation) was performed (usually written in the section of statistical analysis) should be asked.

Please assure the authors do not report “trends” in cases where statistical significance could not be reached. This usually leads to confusion and better be addressed by increasing the sample size or not reporting. In summary, we should scrutinize the results closely and interpret them carefully by checking the abovementioned points.

28.5 Important Points to Evaluate in the Discussion

In general, the discussion of a paper serves as a section to interpret the results and present the meaning of the findings. This section typically includes the following:

- Summary of the main findings of the paper.
- Focus on the main findings of the paper, and avoid generalization to a point where the data presented in the current study cannot support the discussion.
- Correlation of the main findings with previous literature.
- Strengths and weakness of the study design.
- Future directions.
- Conclusion/take-home message.

Each one of these components of the discussion should be evaluated to determine the potential clinical impact of the findings presented in the paper.

First, the summary of the main findings of the study should be comprehensive and set the stage for the discussion about these findings. There should be a balance between simply restating all of the results and omitting findings that might be important for interpretation of the results in a clinical context. It is important for the reviewer to ask whether this balance was achieved and to make sure that findings from the results that could possibly add or negate from the author’s interpretation are not omitted.

The main findings that are presented in the beginning of the discussion are typically correlated with previous findings in the literature. The findings of the current study can agree or disagree with those in the literature, but the reviewer

must assess if this agreement/disagreement favors the author’s interpretation of the results. As a reviewer, it is important to have some baseline knowledge about other studies done on the same topic in order to determine if the author’s line of thought makes sense. It is difficult to have a depth of knowledge of certain topics, especially as a young investigator. Thus, in those cases it can be useful to read the papers that are being referenced by the author in the introduction and discussion to gain a better understanding of the literature. With better knowledge of the topic, it will be become easier to follow the author’s logic in the discussion and will allow the reviewer to properly assess if that logic is appropriate.

The strengths and limitation of a study design play an important role when determining the clinical relevance of the paper and also the potential impact the paper can have on clinical practice. Many of the points that will be mentioned in the strengths and limitations section of the discussion are already mentioned earlier in “Reviewing the Methods of a Clinical Research Paper” section of the chapter. However, it is important to highlight the factors that will most directly influence the potential impact of the findings of the study. The number of patients in a study is an important factor since it directly affects the power of the conclusions made in the study. For example, if treatment A was shown to have a decrease in length of hospital stay by 3 days compared to treatment B, but this was based on a trial of only ten patients, it does not influence clinical practice until a larger study is able to show the same effect. Additionally, mean follow-up time and loss to follow-up are also important factors to consider when evaluating the impact of the findings of a clinical research paper.

Future direction of a study provides information regarding the potential for the findings of the study to influence additional research on the topic. This may be used further in the assessment of the impact of the findings of a paper. If the findings will serve as the basis for many future studies, the influence of the paper may be greater than if it was just an isolated finding with no clear path of additional research that leads to clinical relevance.

28.6 Format for a Review to a Journal

Key components for a good peer review are (1) accepting to review in a timely fashion, (2) completing the review in a timely fashion, and (3) providing a high-quality peer review. There are several components to a review for a journal, and the main sections are the following:

- Summary of the paper
- General Comments
- Specific Comments
- Comments to the Editor

The review should start with a summary of the paper that shows the author that you have a clear understanding of the purpose and results of the paper. This should be followed by general comments that should outline the major points that need to be further addressed in the paper prior to being considered for publication. For example, if there is further clarification needed about what statistical method was used for certain comparisons in order to properly interpret the results, this should be included in the general comments. The next section is the specific comments, which include line-by-line comments and overall comments about each section of the paper. The principles for reviewing each section outlined in this chapter play a big role when compiling the specific comments. The comments in this section of the review include correction of grammatical errors, correction of terminology, and clarification of any points that are confusing (this can be in the form of a question to the author). Finally, a paragraph with comments to the editor should be included in the review. This will serve to provide the editor with the strengths and limitations of the paper and the potential fit of the paper in the journal. This section should also include a decision regarding if the paper should be accepted for publication, reconsidered after revision, or rejected. Different journals may have different guidelines for the reviewer, but in general these are the sections that should be clearly distinguished in a review submitted to a journal.

Fact Box 28.2

Resources with guidelines and further education about reviewing a paper:

- <https://journals.lww.com/jbjsjournal/Pages/Consultant-Reviewer-Guidelines.aspx>
- <https://publicationethics.org/resources/guidelines-new/cope-ethical-guidelines-peer-reviewers>
- <http://senseaboutscience.org/activities/peer-review-the-nuts-and-bolts/>
- <https://publons.com/community/academy/>

Take-Home Message

- Reviewing a paper requires a systematic approach in order to ensure that nothing is overlooked.
- Each section of the paper must be evaluated for certain elements.
- The introduction must provide a clearly reasoning for the motivation of the study, the aims of the study, and the hypotheses of the study.
- The methods must be described in detail such that they can be easily repeated and must be of the highest quality according to different metrics set based on the study design.
- Interpretation of the results should be in the context of clinical relevance, not only statistical significance.
- Finally, the discussion should bring everything together and allow the reviewer to determine the impact that the study would have if published.
- Overall, the guidelines presented in this chapter for evaluation of each section of a paper will serve as a good foundation that reviewers can adapt while reviewing specific papers.

References

1. Brand RA. Writing for clinical orthopaedics and related research. *Clin Orthop Relat Res*. 2008; 466(1):239–47.
2. Higgins J, Green S. *Cochrane handbook for systematic reviews of interventions*. Chichester: John Wiley & Sons; 2008. <https://search.library.wisc.edu/catalog/9910060197402121>.

3. Moher D, Shamseer L, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev*. 2015;4:1. <https://doi.org/10.1186/2046-4053-4-1>.
4. Ramisetty N, Kwon Y, Mohtadi N. Patient-reported outcome measures for hip preservation surgery—a systematic review of the literature. *J Hip Preserv Surg*. 2015;2(1):15–27. <https://doi.org/10.1093/jhps/hnv002>.
5. Schulz KF. Subverting randomization in controlled trials. *JAMA*. 1995;274(18):1456–8. <https://doi.org/10.1001/jama.1995.03530180050029>.
6. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Med*. 2010;8(1):18. <https://doi.org/10.1186/1741-7015-8-18>.
7. Stone AV, Jacobs CA, Luo TD, et al. High degree of variability in reporting of clinical and patient-reported outcomes after hip arthroscopy. *Am J Sports Med*. 2017;46(12):3040–6. <https://doi.org/10.1177/0363546517724743>.
8. Thorborg K, Tijssen M, Habets B, et al. Patient-Reported Outcome (PRO) questionnaires for young to middle-aged adults with hip and groin disability: a systematic review of the clinimetric evidence. *Br J Sports Med*. 2015;49(12):812. <https://doi.org/10.1136/bjsports-2014-094224>.

Part V

How to Perform a Clinical Study: A Case Based Approach

Level 1 Evidence: A Prospective Randomized Controlled Study

29

Seper Ekhtiari, Raman Mundi, Vickas Khanna, and Mohit Bhandari

29.1 Introduction

The documentation and reporting of clinical scenarios have been an important part of medical practice for millennia. The earliest identified example of what could be considered a case series dates to at least 1600 BC. The ancient Egyptian papyrus (*Edwin Smith Papyrus*) describes the presentation and management of 48 different conditions, primarily traumatic injuries [45]. Despite this historical precedent, much of medical literature remained quite rudimentary for most of the intervening time. The application of true scientific rigour to medicine was rarely seen before the 1700s, when Enlightenment ideologies began to permeate into medical practice. It is difficult to establish the “first” controlled clinical trial with certainty, but one of the most likely candidates was a trial conducted by Dr. James Lind in 1747. Lind was a Scottish surgeon who demonstrated through a controlled trial that citrus fruits were effective against scurvy in seafaring sailors [9]. Throughout the eighteenth and nineteenth centuries, the use of quasi-random allocation methods, most notably alternate allocation (i.e. “every other patient”), became increasingly common and important [12].

Perhaps the first example of a true randomized controlled trial (RCT) is Amberson’s 1931 study, in which a group of relatively similar patients were allocated by coin flip to sanocrysin or no treatment for tuberculosis [5]. Another tuberculosis trial, conducted nearly two decades later, likely represents the most important turning point in the history of RCTs in medicine. Sir Austin Bradford Hill was the statistician overseeing a streptomycin trial for the treatment of tuberculosis and was the first to utilize and formalize some core components of the modern RCT, including truly randomized allocation, allocation concealment, and blinding [9].

The use of RCTs in orthopaedic surgery, however, remains limited, partly due to the unique and inherent challenges involved in conducting a surgical RCT [15]. Despite these challenges (discussed below in further detail), RCTs remain the gold standard of evidence-based medicine and should continue to be the ultimate goal when conducting clinical studies. Randomly assigning patients to treatment groups that are concealed from both the patient and the investigator and comparing the different groups minimizes the risk of selection bias and mitigates the effect of some psychological factors that can impact the outcome of a study. Finally, such a trial can help to decide if a new treatment is worse, similar, or better than the existing standard of care. Thus, well-executed RCTs do represent a true pinnacle on the pyramid of evidence-based medicine [4].

S. Ekhtiari · R. Mundi · V. Khanna · M. Bhandari (✉)
Division of Orthopaedic Surgery, McMaster
University, Hamilton, ON, Canada
e-mail: bhandam@mcmaster.ca

29.1.1 Definitions

Many terms describing various aspects of clinical trials are often used interchangeably and at times incorrectly. A detailed understanding of the nomenclature is essential in designing, labelling, evaluating, and interpreting clinical trials. These terms are briefly defined in this section. Akobeng provides an excellent and detailed review of each component [4].

- **Prospective** refers to the *a priori* planning of various stages of the trial. True RCTs are, by definition, prospectively designed as randomization and allocation can only occur prospectively. “Prospective” can apply to the study design, the data collection, and the data analysis. In a study that is fully prospective by design, data is collected for a predefined study, and the data analysis is performed based on a predetermined plan. However, not all publications stemming from an RCT are necessarily fully prospective; prospectively collected data from an RCT can also be analysed retrospectively. Prospective data collection tends to capture more precise and accurate data as the data collection instruments have been selected and/or designed specifically to address the study objectives [19].
- **Randomization** refers to the method by which patients are allocated to the various treatment and/or control groups. Ideally, this should be a truly random method that has no potential for participant or researcher interference. This does *not* include quasi-random methods of allocation such as chart numbers, birth dates, or alternate allocation as often unrecognized bias or trends can influence the allocation. Specific methods of randomization are discussed below.
- **Blinding** (or “masking”) refers to the process by which the treatment that each patient is receiving remains unknown to one or more of five groups of individuals: the patient, the care provider(s), the outcome adjudicator(s), the data collector(s), and the data assessor(s) [42].
- **Controlled** refers to the fact that the treatment(s) of interest is being compared

directly to something else—this may include any or all of the following: a different treatment, a placebo, or no treatment. The term “control group” is used with various meanings both colloquially and in the scientific literature. For the purposes of this chapter, the terms “control” and “control group” will refer to all of the above possibilities.

The following chapter will focus on basic methods for conducting RCTs in orthopaedic surgery. The chapter will be separated into three main sections: (1) planning for an RCT, (2) conducting the RCT, and (3) knowledge dissemination and translation. Each section will contain a detailed discussion of the steps required for conducting a successful RCT. Throughout the chapter, the Fluid Lavage of Open Wounds (FLOW) trial [61] will be used as a case study to demonstrate the practical and real-life application of the topics discussed. The FLOW study was a randomized, multicentre, controlled trial looking at the management of open fractures. Specifically, the study compared irrigation pressures (very low vs. low vs. high pressure) and irrigation

Fact Box 29.1: “Prospective Blinded Randomized Controlled Trial”:

Words with a Meaning

1. **Prospective:** All planning, including data collection instruments and data analysis planning, is completed before the study is initiated.
2. **Blinded:** One or more of the groups of people involved are kept unaware of the group allocations.
3. **Randomized:** A true randomization technique is used to allocate patients to group.
4. **Controlled:** There is a comparison group (may be placebo, no treatment, or other treatment).
5. **Trial:** A clinical study.

Recommended resource: Akobeng AK. Understanding randomised controlled trials. *Arch. Dis. Child.* 2005;90:840–4.

solution (soap vs. normal saline) and their effects on reoperation rate in open fractures. The study randomized a total of 2551 patient, of whom 2447 were included in the final analysis. The investigators found that reoperation rate was similar across irrigation pressures but was significantly lower for normal saline compared to soap [61].

29.2 Planning for a Randomized Controlled Trial

The adage “if you fail to plan, you are planning to fail” certainly applies when it comes to conducting RCTs. There are several crucial steps that need to be taken before embarking on a full-scale RCT. These preliminary steps serve many functions, including setting the foundation for the project, revealing technical and logistical challenges with conducting the study, and verifying that the research question is indeed feasible, novel, and interesting. These steps include defining the research question(s), performing literature reviews, conducting surveys, executing pilot studies, calculating required sample sizes, and assembling the necessary support structures.

29.2.1 Defining the Research Question

An appropriate, clearly defined research question is an indispensable part of conducting research at all levels of evidence. As the foundation of the entire project, this is arguably the most important step of the research process. As Kumar (2005) puts it, “it is like the identification of a destination before undertaking a journey...in the absence of a destination, it is impossible to identify the shortest—or indeed any—route”. There are certain requirements that should be considered in the development of a good research question. Hulley et al. have suggested the use of the acronym FINER, which has become widely used and accepted as important criteria for a good research question [29]. The acronym stands for: feasible, interesting, novel, ethical, and relevant.

Feasibility refers to the practical and logistical aspects of conducting a study. There are various considerations within this component. Sample size is an important consideration: a sample size calculation (discussed later in this chapter) can help to estimate the number of subjects required for adequate statistical power. A review of hospital data can provide a rough estimate of the potential number of subjects that can be expected to be available for recruitment. These numbers should be treated cautiously, and factors such as declining participation, loss to follow-up, ineligibility for inclusion, and random fluctuations in patient volumes should be considered. The number of subjects required/available will also inform the timeframe of the study. It is important to ensure that adequate technical expertise is available for the completion of the project. Multicentre collaboration may be necessary if the subject matter or outcome is highly subspecialized and relatively rare. Funding is an important issue, and a preliminary budget outline can be helpful in projecting the expected costs of all aspects of the study. If external funding (e.g. government research grants) is required, potential funding bodies should be identified early to ensure the project is designed in a way that meets their requirements and mandates. Finally, the scope of the project should be broad enough to answer the research question, but not too broad as to cause confusion and risk losing focus on the underlying question [29].

The research question should be **interesting**. This may sound self-explanatory but can be difficult to accomplish in actual practice. What may seem interesting to a group of surgeons at one institution may be of limited or no relevance to most other surgeons. This does not preclude the execution of the study, but does impact its applicability and generalizability. Discussions with experts in the field and potential funding bodies are good starting points to assess interest in the idea [29]. Ultimately, a large survey of surgeons in the relevant field is an effective way to objectively gauge interest in the research question and assess its potential to change practice [46].

Novelty is another important component of a good research question. This is perhaps the most

Case-Based Example 1: FLOW Survey: Overview

Before embarking on the FLOW study, the investigators conducted an international survey of nearly 1000 orthopaedic surgeons to identify their practice patterns and the likelihood that they would change those practice patterns based on evidence from a large RCT. The vast majority of respondents (94.2%) stated that they would change their practice based on such data [46]. Some important factors in this high level of interest likely include the potential to improve outcomes in a highly vulnerable population, the presence of alternatives that are widely available and relatively inexpensive (high vs. low-pressure irrigation, normal saline vs. soap), and a paucity of pre-existing clinical evidence.

nanced and least universal of the FINER criteria. Hulley et al. (2007) state that “good clinical research contributes new information” [29]. While this is certainly true, the concept of “new information” can be considered in various ways. The research question does not need to be entirely new. A previous research question can be applied to a new population, a new tool or measurement can be used to assess a previous research question, and so on. As well, the importance of replication as part of the scientific method cannot be overstated. A recent analysis of education research found that only 0.13% of publications in the top 100 education journals were dedicated to replicating previous research [37]. Thus, while novelty is important, it can be interpreted in many different ways and should not be taken to mean that every research project has to have an entirely original research question.

While it may seem rather obvious that a research question should be **ethical**, the history of clinical research contains many cautionary tales of unethical research, including some relatively recent examples. The turning point for ethics as a cornerstone of human research occurred following World War II, with the estab-

lishment of the Nuremberg Code. Despite this, the Tuskegee Syphilis Study continued to enrol low-income African-American subjects until 1972. These patients were not told about their syphilis status, and were not offered penicillin, a proven therapy for the disease [39]. Institutional ethics board requirements should be consulted for each specific study, but the Helsinki Declaration provides useful general guiding principles [67]:

- Research with humans should be based on basic science where possible (i.e. lab and/or animal models).
- Research protocols should be reviewed by an independent research ethics board (REB) prior to initiation.
- Informed consent must be obtained.
- The individuals conducting the study should be adequately qualified and trained.
- Risks should not exceed benefits.

Whereas the FINER criteria provide useful general requirements for a good research question, the PICOT format presents a practical way for defining and communicating the research question. This format was first outlined by Richardson et al. as “PICO”: population, intervention, comparator, and outcome [48]. It was later expanded to PICOT to include the time-frame over which the study would be conducted [24]. Like the FINER criteria, the PICOT format is applicable to all levels of evidence. The individual components of the PICOT format are outlined below.

The **population** of interest should be well defined. This is mostly accomplished through clear and well-defined inclusion and exclusion criteria [60]. Important components include basic demographic information (age range, sex, etc.), the nature of the injury/disease (e.g. acute vs. subacute/chronic), and special populations (e.g. paediatric or pregnant patients, etc.). Some components of the target population may need to be considered even though they are not formally part of the inclusion and exclusion criteria. For example, two hypothetical studies conducted in Brazil and China may have identical inclusion/

exclusion criteria, but will inevitably have some inherent differences in the populations they recruit.

The **intervention** of interest is often a new technique, implant, or adjunct that is the primary focus of the study. The opportunity should be taken at this early stage to define the intervention clearly and in detail. All pertinent and applicable details should be recorded. If a new surgical technique is being studied, a detailed description of the operation, often published separately as a technique article, is important. The **comparator** is essentially the control group[s]. Again, as much detail as possible should be included. The type of comparator (e.g. no treatment vs. placebo vs. alternate treatment) should be explicitly stated and clearly defined [62].

The **outcome** is at the heart of why the study is being conducted. What is expected to improve (or not) with this new or different intervention? Generally, in orthopaedics, there are three broad categories of outcomes that are important to evaluate: generic outcomes, condition-specific outcomes, and utility outcomes.

Case-Based Example 2: Excerpt from FLOW Protocol

The FLOW investigators described all intervention arms in great detail in their original study protocol, which was published [62]. For example “...will use sterile technique to inject 80 mL of the clear liquid soap (Castile Soap, Triad Medical Inc. Franklin, Wisconsin—17% concentration in de-ionized water preserved in 90 mL bottles)”. Irrigation pressures were also defined objectively with PSI cut-offs. This level of detail is important for multiple reasons: it allowed the study to be conducted consistently across 41 sites on four different continents, simplifies any future repetition of the study, it made for clear and consistent communication in subsequent publications, and it minimized the effects of varying protocols and practice patterns.

Finally, the planned **timeframe** is important to identify. Again, as with costs, this will not be an exact projection, so some buffer room should be accounted for in case some steps take longer than expected. The timeframe also has significant logistical implications: budgeting for research and administrative staff can vary significantly depending on the timeframe, and the graduation and potential relocation of learners may need to be accounted for.

The FINER and PICOT criteria can be applied to the development of primary and secondary research questions. The primary question should be a single, clearly stated question driven by a hypothesis that forms the basis for the design of the study and the choice of data collection instruments [21]. Secondary research questions are other outcomes related to the intervention. Any results or answers to secondary questions should be considered preliminary rather than definitive [64]. As well, secondary questions should be limited in number to avoid the trap of many comparisons. When many comparisons (i.e. statistical tests) are made, the chances of finding a spurious result increase, unless the appropriate post hoc statistical corrections are performed to account for the multiple comparisons [21].

29.2.2 Literature Review

A detailed description of how to conduct literature reviews, including systematic reviews, can be found later in this book. Briefly, a thorough review of the literature is important to ensure that the research question is novel and that clinical equipoise exists. Clinical equipoise refers to the fact that for a given comparison of treatment arms (e.g. operative vs. nonoperative, drug vs. placebo, operation X vs. operation Y, etc.), a consensus of experts would not consider one arm clearly superior to the other [25]. Clinical equipoise is an important requirement of an RCT, both from an ethical standpoint and in terms of potential clinical applicability. In addition to published literature, it is important to consider sources of unpublished and in-progress literature,

such as trial registries (e.g. www.clinicaltrials.gov), conference abstracts, and online research communities where researchers can post their current works in progress (e.g. www.researchgate.net).

29.2.3 Surveys

A survey of experts in the field is not always a component of preparing for an RCT, but it serves as a valuable tool and should be considered in most cases. A well-conducted survey can provide an objective basis upon which to develop the RCT. It can also be an asset in requesting funding, as it can demonstrate clearly that the question is interesting, relevant, and important. It is important to design and conduct the survey in a fashion that maximizes the survey's validity and response rate. Sprague et al. provide a list of 12 principles for conducting a survey in orthopaedic surgery, along with some important tips and strategies [57].

Case-Based Example 3: FLOW Survey: A Closer Look

Prior to conducting the FLOW study, the investigators surveyed an international group of surgeons. The survey was developed using focus groups, previous literature and key experts. The questionnaire was then pre-tested with an independent group of four orthopaedic surgeons to assess face and content validity. A sample size analysis was performed for the survey to estimate the number of participants required to achieve sufficient statistical power. This analysis estimated that at least 930 questionnaires needed to be distributed, based on a 70% response rate, and 650 completed surveys needed. The investigators distributed 1764 surveys by mail and in-person at a trauma course. Ultimately, there was a 56% response rate,

with 984 completed surveys. Had the investigators distributed only 930 questionnaires with the same response rate, they would have had an insufficient number of completed responses (521) [46]. This demonstrates the importance of overshooting sample size targets where possible.

Case-Based Example 4: FLOW Pilot Study

The FLOW pilot study randomized 111 patients and followed them for 1 year. Two main issues were identified and addressed before the full-scale trial was conducted. First, three of nine sites were found to have below-target compliance rates. These sites received additional training prior to participation in the FLOW RCT. In addition, the loss to follow-up rate of nearly 20% was identified as an area for improvement. The investigators outlined five strategies that they used in the definitive trial in order to improve retention [7].

Fact Box 29.2: Defining the Research Question

PICOT and FINER are useful acronyms to keep in mind when defining the research question. FINER contains the overarching elements that should be included in a research question, whereas PICOT provides a practical method for communicating the question.

FINER: Feasible, Interesting, Novel, Ethical, Relevant

PICOT: Population, Intervention, Controls, Outcomes, Timelines

Recommended resource: Farrugia P, Petrisor BA, Farrokhyar F, Bhandari M. Practical tips for surgical research: Research questions, hypotheses and objectives. *Can. J. Surg.* 2009;53:278–81.

29.2.4 Pilot Studies

Pilot studies are smaller, shorter versions of the trial that may eventually be conducted. The primary purpose of a pilot study is to evaluate the feasibility and logistical practicality of the proposed study. The purpose of a pilot study is *not* to test the hypothesis. In fact, results from pilot studies should not be interpreted in the same way as results from full-scale clinical studies. The main focus during the pilot process is to identify and resolve challenges in recruitment, randomization, blinding, and retention [7]. In most cases, data from a pilot study should not be included in the eventual analysis of RCT data. This is because methodological changes are often made based on the pilot study, which adds a source of uncontrollable variability to the data [34].

29.2.5 Sample Size Analysis

A thorough review of sample size analysis can be found in Chap. 20 of this book and thus will not be discussed in detail here. Briefly, it is important to perform sample size calculations at each stage of the process, including survey administration, pilot study, and full-scale RCT. Care should be taken not to be overly optimistic about recruitment, eligibility, and retention rates [49]. In addition, the treatment effect tends to be overestimated, particularly if the expected effect size is based on previous small studies [1]. Thus, where possible, the final sample size target should be greater than the calculated sample size by 10–40% [63].

29.2.6 Support Structures

It is important to identify, recruit, and train the necessary personnel required to conduct the RCT before the trial has begun. Pilot studies can be very helpful in identifying strengths and limitations of the research team and areas

where additional support may be necessary. It is important to consider the entire spectrum of support staff that may be necessary: volunteers, research assistants, financial experts, grants administrators, and statisticians. The need for these various roles may fluctuate over the different stages of the trial, and commitments and contracts should be tailored as such with some flexibility to allow for unexpected circumstances.

29.3 Conducting a Randomized Controlled Trial

Conducting an RCT is a long process with many stages, some of which occur simultaneously. An awareness of the trial's stage of progress and active preparation for upcoming stages are an important factor in the smooth operation of an RCT. The main steps of conducting an RCT are as follows: REB approval, trial registration, patient recruitment, randomization, allocation concealment, blinding, control and intervention implementation, follow-up, and statistical analysis. These steps are discussed in the below section.

Case-Based Example 5: FLOW Study Support Team

The FLOW study employed one full time research coordinator and one full time undergraduate research assistant for the entirety of the trial. In addition, a project manager worked about half time throughout the project to oversee the project and manage finances. A data manager and a statistician were available for the entirety of the trial, but the early phases required mostly data management whereas towards the end of the trial, the statistician was required nearly full time. Finally, a grants administrator helped to prepare and submit grants for funding opportunities.

29.3.1 REB Approval and Trial Registration

Prior to conducting an RCT, institutional REB approval should be obtained. In addition, any necessary funding should be applied for and secured. Both processes will generally require a detailed protocol of the study, including the literature review, objectives, methods, hypotheses, timelines, and detailed budget. Chapter 8 in this book provides a detailed overview of how to prepare a study protocol. Once the protocol is complete, it should be registered with the appropriate trial registry. In North America, this is most commonly www.clinicaltrials.gov and in the European Union www.clinicaltrialsregister.eu. Prospective registration of all clinical trials is important because it ensures transparency, prevents duplication, reduces publication bias, and lessens the likelihood of selective reporting [6].

29.3.2 Patient Recruitment

A useful way to think about patient recruitment is to think about the “who, what, where, when, why, and how”. A detailed recruitment plan is important to have at the protocol stage, as adequate sample size is integral to RCT success. An unsuccessful recruitment strategy can serve as a bottleneck that slows down or altogether halts an otherwise well-planned trial.

Who will recruit the subjects? It is important to identify and train the individuals who will be recruiting patients. Care should be taken to ensure that these individuals are not those directly responsible for the care of the patient. Involving the patient’s care team (physicians, nurses, etc.) can be an important starting point as there will already be an established rapport. However, it is preferable that the role of the patient’s care team be limited to informing the patient about the research project and evaluating whether the patient would be willing to meet with a member of the research team [55]. An independent member of the research team should then obtain informed consent, and it should be made clear that the decision to not participate in the trial will not unduly affect the patient’s care. Resident

physicians are often the first members of the orthopaedic surgery team to meet a patient and thus can play an important role in maximizing recruitment by identifying eligible patients. If a trial is time-sensitive (e.g. time from presentation to operating room is a primary outcome or an independent variable), it may be necessary to train triage and other emergency department staff to identify and flag eligible patients for recruitment.

What will the recruitment process entail? A detailed informed consent form, which includes a patient copy, is necessary but not sufficient in patient recruitment. It is important that the individuals obtaining informed consent are well informed about the purpose of the study, the various treatment arms, and the potential risks and benefits of each. Copies of the informed consent form should be readily accessible and available to members of the research team at key patient encounter locations (e.g. emergency department, clinic, etc.). As well, if any incentives are being provided for participation (e.g. gift cards, etc.), these should also be available in advance in a secure location.

Where and when will the recruitment take place? Depending on the details of the RCT, recruitment may take place in the emergency department, on inpatient wards, or in outpatient clinics. In the case of multicentre studies, the ideal recruitment location may vary between the different sites. Thus, consultation with staff at each site is important in the planning and pilot stages to ensure that the protocol is consistent but adaptable to each site. It is also important to define *a priori* the “hours” of the trial. This is particularly applicable if the trial is focused on emergent and/or traumatic conditions, as they may present at any hour. Realistic goals should be set depending on the availability of research support staff. For example, the HIP ATTACK trial is an ongoing, international RCT evaluating accelerated (within 6 h of diagnosis) vs. standard care for hip fracture patients. For the enrolment of each patient, a research staff is required to enrol and randomize the patient in a timely manner to achieve the target time for patients randomized to accelerated treatment. Thus, in this trial, subjects are only recruited during daytime

working hours, although each recruitment site may have differing definitions of working hours depending on staff availability [28].

Though patient recruitment in surgical trials most often occurs in person, there are multiple other methods for patient recruitment. These include media outlets, physician referrals, cold calls/mailings, and online recruitment [63]. Previous literature has identified physician referrals and online recruitment as the most cost-effective strategies [63]. Local and institutional regulations should be consulted before public recruitment strategies are considered.

How will the recruitment be performed and why? There are multiple different recruitment strategies, and the most appropriate method should be chosen to fit the study design. These strategies include [63]:

1. All patients recruited simultaneously and begin trial at the same time.
2. Patients enter trial in a “batched” fashion.
3. Continuous recruitment until target sample size is reached.
4. Continuous recruitment until a target end date is reached.

Each strategy is better suited for certain study designs than others. For example, recruiting all patients and beginning a trial simultaneously works best for nonoperative trials, where many patients can be instructed to begin a therapy on or about the same date. Batch enrolment can work well for common, elective procedures, such as total joint arthroplasty. For example, all eligible patients undergoing a total joint arthroplasty in each week can be recruited and enrolled in the trial, and this process can be repeated as necessary. Continuous recruitment is the most common recruitment strategy in surgical trials, particularly when it comes to trauma. For most RCTs, recruitment is continued until the target sample size is reached. If there is a clear rationale to stop recruitment on a specific date (e.g. seasonal conditions/injuries), this strategy may be considered.

Patient recruitment difficulties may be encountered with various aspects of the RCT. These include the protocol itself, staff- or site-specific issues, surgeon-related issues, and patient-related

recruitment issues. Thoma et al. provide a detailed guide to troubleshooting each type of recruitment challenge [63].

29.3.3 Randomization

Randomization, as defined above, is essentially the process by which subjects are allocated to the various groups of an RCT. It should be noted that methods such as the use of chart numbers, birth dates, or alternating allocation are not truly random and thus considered “quasi-randomization”; they should be avoided if possible [17]. Even truly random methods that are manually performed (e.g. coin flips, dice throwing, etc.) are vulnerable to user error and/or interference. The use of secure, computer-generated randomization software is usually considered the best method for random allocation [17].

There are at least four different types of random allocation, all of which may be acceptable for an RCT depending on the specific scenario:

- **Simple randomization** is most commonly performed using a random table number or computer-generated randomization software. This is the easiest randomization technique to employ, but should generally be reserved for large clinical trials. In smaller clinical trials, this method could result in groups that are unbalanced in terms of total numbers and/or baseline characteristics [59].
- **Block randomization** is useful in smaller trials to ensure similar allocation between groups. The “block size” refers to the number of patients that are randomized at a time and should be a multiple of the number of treatment groups (e.g. if two treatment groups exist, block size can be 4, 6, 8, etc.). Subjects in these small blocks are then randomized in a balanced fashion to the treatment groups. For example, in a study with two treatment groups (A and B) and a block size of four, the first four subjects can be randomized in any of the following combinations: AABB, ABAB, BBAA, ABBA, BAAB, and BABA. One of these six combinations is then selected at random and applied to the first four patients. The

process is then repeated for each subsequent group of four subjects [4, 59]. The block sizes can be constant or variable for each allocation.

- **Stratified randomization** is a strategy that attempts to minimize the likelihood of having significantly different covariates between patient groups. In this type of randomization, patients are grouped based on select, predetermined prognostic factors before being randomized. For example, in a study on the wound complications of total joint arthroplasty, the presence of diabetes mellitus (DM) is an important prognostic factor. Thus, patients could be stratified based on DM status before being randomized to a treatment arm. Though this type of randomization is appealing due to its potential to control for covariates, it can become complicated if many covariates need to be controlled. As well, this type of randomization requires that all participants be identified before randomization occurs [59]. While this may be well suited for some elective procedures (e.g. by identifying patients on a waiting list for the same operation), it would not be applicable for any emergent or trauma-related RCT.
- **Covariate-adaptive randomization** is a strategy that attempts to address some of the limitations of stratified randomization. Essentially, subjects are enrolled sequentially in real time, and the covariates of interest (e.g. age, comorbidities, etc.) are entered for each patient. Covariate-adaptive randomization attempts to randomize new subjects into groups by considering the previously randomized subjects and correcting for any imbalances between groups [59].

29.3.4 Allocation Concealment

Allocation concealment refers to a situation in which the individual enrolling the patients and performing the randomization does not know which group the next subject will be randomized to. Previous literature has shown that effect sizes

are overestimated by 41% in trials that do not perform allocation concealment [52]. Traditionally, a common technique for allocation concealment was the use of opaque, sealed envelopes, which a member of the research team would open for each patient at the time of randomization. The *Cochrane Handbook of Systematic Reviews of Interventions* designates the use of sealed opaque envelopes as carrying a “low risk of bias” [27]. Empirical data, however, suggests that trials utilizing sealed opaque envelopes are prone to tampering [33] and are more likely to demonstrate statistically significant results compared to those using distance randomization [26]. Distance randomization occurs when the randomization process is completely removed from the control of the individual(s) enrolling the subjects. Usually, this occurs either via telephone or the Internet, whereby the research team member calls or logs into a secure, centralized service that then randomizes the subject and records their allocation. Distance randomization, particularly through third-party, secure websites, is the preferred method for randomizing subjects in today’s RCTs.

29.3.5 Blinding

Blinding is a distinct concept from allocation concealment; it refers to the process by which one or more of the following groups are kept unaware of the subject’s assigned treatment group: (1) the subjects, (2) the caregivers, (3) the outcome adjudicators, (4) the data collectors, and (5) the data analysts. Depending on the study design, any or all of these five groups may be blinded [31]. According to the Consolidated Standards of Reporting Trials (CONSORT) statement, the commonly used terms “single blinded”, “double blinded”, and “triple blinded” should be avoided, as they are ambiguous and uninformative. Rather, investigators should specifically state which, if any, of the five groups outlined above were blinded [43]. As well, it is important to clarify if the same person or group fulfilled more than one of the roles. For example, in a surgical trial, the primary surgeon can potentially be

the caregiver, the outcome adjudicator, and the data collector.

As a rule, as many groups as possible should be blinded to the group allocation. In practice, particularly in the context of surgical trials, this is not always possible. For example, it is impossible to blind the surgeon to an operative intervention [44]. Inability to blind one or more groups, however, should not automatically result in a complete lack of blinding. A systematic review of orthopaedic trauma RCTs found that less than 10% of RCTs reported blinding outcome adjudicators. Interestingly, up to 96% of outcome adjudicators could have been blinded with simple methods not requiring major changes to the RCT protocol [30]. A more recent review of all surgical literature found that 52% of RCTs blinded the outcome adjudicators, compared to 67% of RCTs, which could have done so [56]. This review, however, was not systematically performed and only examined the ten highest-impact journals in medicine and surgery and thus may paint an overly optimistic picture.

Blinding is particularly important in orthopaedic RCTs, as many outcome measures are focused on patient-reported levels of function, pain, and quality of life. Thus, creative strategies may be required to blind each group of individuals. Blinding subjects is a simple task in some medical and procedural trials. For example, a trial of intra-articular knee injections comparing corticosteroids to placebo was performed with the subjects blinded [47]. By having the syringes prepared by an independent healthcare provider, the caregivers could have also been blinded in this trial. Similar strategies can be used for other nonoperative trials that do not require a visible construct such as a specific sling or brace type. Blinding of subjects can also be relatively straightforward for trials comparing two different types of operations, particularly if both operations use similar surgical approaches. In trials comparing operative to nonoperative management, however, blinding becomes more challenging. Blinding of subjects can be performed through sham surgery, which also provides an excellent control for the placebo effect of having undergone an operation [44]. Sham surgery is

discussed in greater detail below, but this is essentially the only way to truly blind subjects in an operative vs. nonoperative trial.

Blinding surgeons in an operative trial is almost always impossible or highly impractical unless the intervention is an adjunct that can be easily masked (e.g. intraoperative local anaesthetic injection). Blinding of the remaining three groups, however, can and should be done in many cases. Blinding of outcome adjudicators and data collectors can be performed in several different ways. Surgical scars can be concealed with patient clothing or large dressings [38]. Even radiographs showing different implants can sometimes be masked using creative digital alteration, as demonstrated by Karanicolas et al. [31]. Radiographs and other imaging modalities can be de-identified prior to being provided to the outcome adjudicator. In the case of patient-reported outcomes, an extra, independent individual may be required to collect data from the patient, de-identify it, and then provide it to the data assessment team. Blinding of the data assessment team is perhaps the simplest form of blinding to achieve, only requiring that data is coded in such a way that allows comparison between groups but does not allow the data analyst to decipher each subject's allocation. Despite this, no orthopaedic RCTs published between 1988 and 2000 specified their use of this technique [8].

Case-Based Example 6: Randomization, Allocation Concealment, and Blinding in the FLOW Study

The FLOW study randomized subjects using a custom-built web-based randomization system using variable block randomization. Given that this was a 'distance randomization' technique, allocation concealment was ensured. Patients, outcome adjudicators, and data analysts were all blinded to subject allocation. In addition, a central adjudication committee, also blinded to group allocation, assessed all subjects for eligibility before or shortly after randomization.

29.3.6 Control Groups

An appropriate choice of control group(s) is of paramount importance in the success and impact of an RCT. There are many different types of control groups that can be used in surgical trials. In general terms, these include no treatment controls, placebo controls, and active treatment controls. From a strictly scientific standpoint, the best control group is a placebo group—i.e. a group that receives no active intervention, but is blinded to this fact. This type of control can be implemented with relative ease in trials of oral or injected medications. As discussed earlier, placebo groups in the context of operative trials essentially amount to sham surgery. A recent systematic review found six orthopaedic RCTs that utilized sham surgery. Interestingly, all six studies found that sham surgery was as effective as therapeutic surgery. The most recent of these six trials was published in 2012 [36].

Thus, sham surgery is clearly possible in orthopaedic RCTs and has the potential to provide very useful information. There are, however, significant ethical concerns with randomizing patients to receive sham surgery. Patients are exposed to risk, with no therapeutic intervention, which may violate the ethical principles of non-maleficence and beneficence [44]. On the other hand, the argument could be made that by revealing some procedures to be no better than placebo, future patients can avoid undertaking unnecessary risks. This does not, however, account for the fact that performing sham surgery also threatens the trust in the doctor-patient relationship because sham surgery status must be kept blinded even at follow-up. Overall, sham surgery may be appropriate in select cases, particularly when there is true clinical equipoise about the utility of a surgical intervention and when that intervention is minimally invasive, low risk, and common enough to confer the results with significant potential impact [44]. Extra attention should be paid to the informed consent process to ensure that patients truly understand the rationale, logistics, and implications of sham surgery.

For most orthopaedic trials, active treatment and/or no treatment is the most commonly used

and most appropriate choices for control groups. The decision between the inclusion of one or both groups depends on the pre-existing evidence on the current standard of care. If there exists true clinical equipoise about the effectiveness of a given operation, for example, it would be reasonable to randomize patients to either operative or conservative treatment and compare their outcomes. However, if there is a well-established treatment that is known to be effective, it would be unethical to withhold that treatment altogether. Thus, the new intervention should be compared to the standard of care [29]. It can sometimes be difficult to determine whether it would be ethical to withhold the current standard of care. A treatment may have become established as the “standard of care” based on little or no high-quality evidence. Legally, a determination of “standard of care” is often made based on the practice patterns of similar physicians. Thus, even if scientific evidence is lacking, it may be legally perilous to withhold a commonly practiced standard of care [40].

29.3.7 Follow-Up

The intended length of follow-up should be decided *a priori* (see the discussion on PICOT format, Sect. 29.2). Of course, follow-up can then be extended beyond that timeframe, but the target follow-up is a decision made based on the known or expected natural history of the disease or injury. Arrangements should be made in advance to minimize loss to follow-up, such as asking patients if they plan to relocate within the study time period and obtaining multiple alternate contacts for each subject. All loss to follow-up should be carefully recorded, including the reason for loss (e.g. loss of contact, death, withdrawal of consent, etc.). As well, all adverse events, even those fully unrelated to the intervention, should be carefully tracked.

29.3.8 Statistical Analysis

Part III of this book contains a basic but thorough approach to statistics. Therefore, only concepts

specifically related to analysis of RCT data will be discussed here. It is recommended that a statistician, or at least an individual with formal statistics training, be consulted throughout all steps of data analysis for an RCT. The details of these analyses are beyond the scope of this chapter, but are often complex and require the appropriate expertise. However, two important concepts that should be understood for anyone involved in conducting RCTs are the intention-to-treat principle and non-inferiority analysis.

The intention-to-treat principle is the technique of analysing all patients randomized to a given group together, regardless of if they went on to receive the appropriate treatment or complete the study [4]. This simulates the real-life limitations of clinical practice—patients stop attending appointments, relocate, change their minds, are misdiagnosed, or deviate from prescribed treatments. By excluding these patients from the primary analysis, the results would demonstrate an optimistic “best-case scenario” rather than results that can be realistically expected. Thus, the primary statistical analysis, where possible, should be based on the intention-to-treat principle [43]. This concept can often be applied to operative trials [14] but, if applied incorrectly, can also produce confusing and nonsensical results. Malavolta et al. demonstrate this with a clear example (see Fact Box 29.3).

Clearly, care should be taken to apply the intention-to-treat principle appropriately and to data that is amenable to such analysis. Where necessary, it is permissible to restrict some of the secondary analyses to those who completed the study protocol (i.e. protocol analysis).

Non-inferiority refers to a study design that simply seeks to show that the new intervention is “no worse” than the current standard of treatment. “No worse” can be defined in various ways, including a statistically significant difference, a minimal clinically important difference (MCID), and more. This contrasts with a “superiority trial”, which may be thought of as the classic RCT. In a superiority trial, the goal is to demonstrate that the new intervention is superior to the current standard. Non-

Fact Box 29.3: The Challenge of the Intention-to-Treat Principle in Surgical RCTs

For example, a patient drawn for non-surgical treatment who, for any reason, then undergoes surgical treatment should, according to the intention-to-treat principle, still be analyzed as a non-surgical case. If this patient happens to present infection at the surgical site, the results from such a study would show “occurrence of infection of the surgical site” as a “complication from non-surgical treatment” [38].

Recommended resource: Malavolta EA, Demange MK, Gobbi RG, Imamura M, Fregni F. Randomized Controlled Clinical Trials in Orthopedics: Difficulties and Limitations. *Rev Bras Ortop.* 2011;46:452–9.

inferiority analysis may be desirable as a first step if the sample sizes required for a superiority trial are unrealistic. For example, in comparing a new intervention with the current standard of care, the incremental effect size difference is likely to be relatively small, and thus the sample size required for the RCT would be very large. Alternatively, if the goal is to demonstrate that the new intervention is “no worse” in a secondary outcome, but not in the primary outcome, a non-inferiority trial can be performed [65].

A final note on statistical analysis in RCTs: according to the CONSORT statement, baseline differences among treatment groups should not be tested for statistical significance. The reasoning behind this assertion is as follows: assuming that the trial has been conducted with rigorous and appropriate methodology and that the randomization technique selected is appropriate for the trial, any baseline differences between the groups are necessarily due to chance. Thus, a test of statistical significance, which assesses the likelihood that these differences were due to chance, is redundant and unnecessary; in fact, its results may be misleading [43].

29.4 Limitations of Randomized Controlled Trials

Randomized controlled trials do present certain challenges when it comes to surgical disciplines and particularly when an operative intervention is included as one of the treatment arms. Many of these concepts are discussed above; below is a summary of the major limitations of RCTs for assessing operative interventions. Certain specific study designs, such as crossover designs, are simply not an option with surgical intervention [38]. As well, while “placebo” (i.e. sham surgery) trials have been conducted and have provided some very important evidence (e.g. in the case of knee arthroscopy), they do present ethical and public relations challenges [41]. The intention-to-treat principle can be difficult and confusing to apply if the trial is comparing an operative treatment group to other treatment groups [38]. As well, surgical technique and expertise vary between surgeons and institutions. An interesting example of how this may skew the results of an otherwise well-conducted trial is in the case of a large, multicentre RCT of displaced intracapsular hip fracture management: consultant surgeons were nearly twice as likely to be present when a patient was allocated to total hip arthroplasty as opposed to fixation or hemiarthroplasty [32].

Another limitation of RCTs may be that their reputation precedes them—in other words, the results of RCTs are sometimes interpreted with a high degree of confidence based purely on the study design. In reality, however, between one-third and one-half of all orthopaedic RCTs are in fact underpowered to detect an adequate effect size [2, 22]. Undoubtedly contributing to this concerning trend is the fact that less than 10% of RCTs published in the highest-impact orthopaedic journals prospectively calculate the adequate sample size required to achieve sufficient power [22, 35]. Thus, the risk of Type II error (i.e. accepting the null hypothesis when there truly is a significant difference) is significant. Perhaps not surprisingly, Lochner et al. found that the rate

of Type II error for primary outcomes in orthopaedic RCTs was over 90% [35].

A challenge unique to surgical RCTs is the issue of surgeon skill and experience. Difficult to quantify, variation in surgical skills can certainly have an impact on the outcomes of an RCT. This is particularly true in cases where the surgeon involved is much more comfortable with one or the other of the treatment arms [50]. The issue of a learning curve for a surgeon performing a new procedure is also a well-established phenomenon, with a certain number of cases required before a steady state is reached [23]. Thus, the use of an expertise-based design for orthopaedic RCTs has increased in recent years, though they remain relatively uncommon overall [16]. Finally, in a field where technology is always rapidly evolving and innovation is constant, the pace of clinical research can at times be too slow to keep up with and evaluate these changes [11].

Fact Box 29.4: Limitations of RCTs in Orthopaedic Surgery

Conducting an RCT in orthopaedic surgery presents a number of unique challenges. Some of these are immutable, while others can be managed to some extent with careful planning. Common examples of these challenges include the use of:

- Certain study designs (e.g. crossover designs)
- True “placebo” arms
- The intention-to-treat analysis
- Blinding (especially surgeons and patients)

Recommended resource: Malavolta EA, Demange MK, Gobbi RG, Imamura M, Fregni F. Randomized Controlled Clinical Trials in Orthopedics: Difficulties and Limitations. *Rev Bras Ortop.* 2011;46:452–9.

29.5 Knowledge Dissemination

Knowledge dissemination is an important part of the scientific process and comes in many forms. Publication in scientific journals is often seen as the logical and desirable end goal of conducting research. This view, however, is increasingly recognized as being too narrow [13]. While scientific journals remain an important pillar of communication within the research community, the results of important scientific studies need to be communicated beyond this silo. The general public and local, regional, and national governing bodies are important target audiences that should be informed about the scientific progress that is being made.

This is important for many reasons. First, much of medical research is funded by government agencies and, thus, public taxpayer money. Furthermore, while publication in scientific journals has the potential to change physician practice patterns, it lacks the platform required to address systemic and funding barriers. For example, if an RCT demonstrates that a certain operation is better and more cost-effective than the current standard of care, but the government continues to incentivize hospitals to perform the standard of care operation, it would be very difficult for any individual surgeon to make the transition to the new technique. Finally, a scientifically literate public has the knowledge and tools to make informed decisions about personal and societal level issues [18].

Of course, publication within peer-reviewed scientific journals remains one important component of communicating the results of an RCT. Various sections within this book address the process of preparing a scientific manuscript. With specific regard to RCTs, the CONSORT statement is an important document that should be reviewed regularly prior to, during, and after the entire RCT process. This statement contains a detailed road map to accurately and consistently communicate the results of an RCT [43]. These guidelines should be adhered to when submitting a RCT manuscript to a scientific, peer-reviewed journal.

Traditional media outlets (e.g. print, radio, television) represent a direct route for communicating the results of scientific studies to the public. An examination of scientific research covered in the news media, by Selvaraj et al., found that the news was more likely to cover observational studies rather than RCTs and that the studies covered were of relatively low quality (even among observational studies) [54]. Thus, it is important for scientists to play an active role in interacting with media outlets, to help accurately and effectively convey the take-away messages of important studies. Press releases play an important role in the flow of knowledge from the scientific community to the journalistic community. About half of all scientific stories appearing in news media originate from a press release [66], and 45% of news stories use the press release as their sole source of information when reporting on the study [53]. Thus, it is important to ensure that press releases are accurate, objective, easy-to-understand, and enthusiastic but not sensationalistic.

More recently, social media outlets have become an important vehicle of communication in almost every sphere of life, including business [51], marketing [3], and politics [58]. The monthly audience for Facebook, Twitter, and online blogging websites have far surpassed the readership of even the most popular forms of print media [10]. Sharing the progress and results of scientific studies through social media can help to engage the public and to raise the profile of a given study in the scientific community. Highly tweeted journal articles are up to 11 times more likely to become highly cited than those without social media coverage [20]. Researchers looking to engage with social media should do so with knowledge and awareness of its intricacies. It is very difficult, if not impossible, to boil down a 3000-word RCT manuscript into 140 characters with the intended message and necessary level of nuance. Social media experts or training may be sought out to ensure that these new outlets are used to their best effects without compromising or misrepresenting the scientific process.

29.6 Conclusion

Randomized controlled trials are generally regarded as the gold standard of scientific evidence, because they are designed to control for as many sources of bias as possible. In planning an RCT, the first step is to come up with a good research question based on the FINER and PICOT criteria. Following this, surveys and pilot studies can help to identify potential issues and challenges with the RCT. Once a protocol has been finalized and the appropriate support teams assembled, patients can be recruited. Recruitment may occur sequentially, in “batches” or all at once. Subjects are then randomly allocated to one of two or more groups using simple randomization, block randomization, stratified randomization, or covariate adaptive randomization. It is important that the allocation is concealed from the individual enrolling subjects. Blinding is desirable, as it limits the effect of suggested or expected outcomes. Blinding can be performed for any or all of five groups of individuals: (1) the subjects, (2) the care providers, (3) the outcome adjudicator(s), (4) the data collector(s), and (5) the data analyst(s). Choice of control group depends on ethical considerations and the presence of a current standard of care and can include no treatment controls, placebo controls, and/or active treatment controls. Try to anticipate and pre-empt issues with loss to follow-up. Involve expert help for the statistical analysis process. Finally, get the word out—to your colleagues, government agencies, the public, and any other important stakeholders.

All attrition and adverse events during follow-up should be documented, even if unrelated to the intervention in question. During the statistical analysis process, it is important to understand the concepts of intention-to-treat and superiority vs. non-inferiority analyses. Baseline differences between groups should not be subjected to tests of statistical significance. Orthopaedic RCTs

face some unique challenges, such as the difficulty with blinding surgeons and sometimes patients, the application of the intention-to-treat analysis, variability in expertise and technique between surgeons, and difficulty keeping up with the pace of technological advances. After an RCT, knowledge can be disseminated in many ways, including through publications in peer-reviewed scientific journal, press releases, traditional media outlets, and social media.

Take-Home Message

- Randomized controlled trials are the gold standard of scientific evidence.
- To conduct an RCT, careful planning is a must.
- Review the CONSORT guidelines as well as definitions of key terms, including blinded, randomized, controlled, and allocation concealment.
- If there is blinding, simply state who was blinded—do not use terms like “double-blind” or “triple-blind” without further elaboration.
- Start with a clear, concise, and well-defined research question—do not rush past this step, and remember the acronyms FINER and PICOT.
- Surveying experts in the field and conducting a pilot RCT can help to identify unexpected barriers, test potential solutions, and aid sample size calculations.
- Plan and budget for every step of the process, from clipboards to management staff.
- Apply for REB approval early, and register the trial on the appropriate publicly available registry before starting.
- Ensure that patient recruitment is performed in a way that is effective but not coercive or disruptive to patient care.
- Use a truly random allocation method, preferably a secure computer-generated software.
- Blind as many groups as possible—this may require some creativity, but is often more feasible than would appear at first glance.

Appendix: Useful Inexpensive Resources

Title	Link	Description
CONSORT Statement and Website	http://www.consort-statement.org	Set of evidence-based requirements for accurate and consistent reporting of RCTs
GraphPad	http://www.graphpad.com	A user-friendly website with many free resources including statistical guides and calculators that can be used for simple statistical operations
SurveyMonkey	http://www.surveymonkey.com	An intuitive survey platform that allows design and distribution of visually attractive, user-friendly surveys. Most accounts are free or relatively inexpensive. Note that not all accounts are compliant with health information privacy legislation
DSS Research Knowledge Center	http://www.dssresearch.com/KnowledgeCenter.aspx	A clinical research website with many free resources, including webinars and a free and intuitive sample size calculator
ClinicalTrials and EU Clinical Trials Register	http://www.clinicaltrials.gov http://www.clinicaltrialsregister.eu	Large international trial registries where new trials can be registered, and ongoing trials can be searched
OxMaR	http://www.ccmp.ox.ac.uk/oxmar	A free, open-source, randomization software developed by the Nuffield Department of Clinical Medicine

References

- Abdulatif M, Mukhtar A, Obayah G, Hardman JG. Pitfalls in reporting sample size calculation in randomized controlled trials published in leading anaesthesia journals: a systematic review. *Br J Anaesth*. 2015;115(5):699–707. <https://doi.org/10.1093/bja/aev166>.
- Abdullah L, Davis DE, Fabricant PD, Baldwin K, Namdari S. Is there truly “no significant difference”? Underpowered randomized controlled trials in the orthopaedic literature. *J Bone Joint Surg Am*. 2014;97(24):2068–73. <https://doi.org/10.2106/JBJS.O.00012>.
- Akar E, Topçu B. An examination of the factors influencing consumers’ attitudes toward social media marketing. *J Internet Commer*. 2011;10(1):35–67. <https://doi.org/10.1080/15332861.2011.558456>.
- Akobeng AK. Understanding randomised controlled trials. *Arch Dis Child*. 2005;90(8):840–4. <https://doi.org/10.1136/adc.2004.058222>.
- Amberson J, McMahon B, Pinner M. A clinical trial of sanocrysin in pulmonary tuberculosis. *Am Rev Tuberc*. 1931;24:401–35.
- Aslam A, Imanullah S, Asim M, El-Menyar A. Registration of clinical trials: is it really needed? *N Am J Med Sci*. 2013;5(12):713–5. <https://doi.org/10.4103/1947-2714.123266>.
- Bhandari M, Guyatt G, Jeray K, et al. Fluid Lavage of Open Wounds (FLOW): a multicenter, blinded, factorial pilot trial comparing alternative irrigating solutions and pressures in patients with open fractures. *J Trauma*. 2011;71(3):596–606. <https://doi.org/10.1097/Ta.0b013e3181f6f2e8>.
- Bhandari M, Richards RR, Sprague S, Schemitsch EH. The quality of reporting of randomized trials in the journal of bone and joint surgery from 1988 through 2000. *J Bone Joint Surg Am*. 2002;84-A(3):388–96.
- Bhatt A. Evolution of clinical research: a history before and beyond James Lind. *Perspect Clin Res*. 2010;1(1):6–10. <https://doi.org/10.4103/2229-3485.103599>.
- Bik HM, Goldstein MC. An introduction to social media for scientists. *PLoS Biol*. 2013;11(4):e1001535. <https://doi.org/10.1371/journal.pbio.1001535>.
- Bothwell LE, Greene JA, Podolsky SH, Jones DS. Assessing the gold standard—lessons from the history of RCTs. *N Engl J Med*. 2016;374(22):2175–81. <https://doi.org/10.1056/NEJMms1604593>.
- Bothwell LE, Podolsky SH. The emergence of the randomized, controlled trial. *N Engl J Med*. 2016;375(6):501–4. <https://doi.org/10.1056/NEJMp1604635>.
- Brownell SE, Price JV, Steinman L. Science communication to the general public: why we need to teach undergraduate and graduate students this skill as part of their formal scientific training. *J Undergrad Neurosci Educ*. 2013;12(1):E6–E10. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3852879&tool=pmcentrez&rendertype=abstract>.
- Bubbar V, Kreder H. The intention-to-treat principle: a primer for the orthopaedic surgeon. *J Bone Joint Surg Am*. 2006;88(9):2097–9.
- Campbell AJ, Bagley A, Van Heest A, James MA. Challenges of randomized controlled surgical trials. *Orthop Clin North Am*. 2010;41(2):145–55. <https://doi.org/10.1016/j.ocl.2009.11.001>.
- Cook JA, Elders A, Boachie C, et al. A systematic review of the use of an expertise-based randomised

- controlled trial design. *Trials*. 2015;16(1):241. <https://doi.org/10.1186/s13063-015-0739-5>.
17. Dettori J. The random allocation process: two things you need to know. *Evid Based Spine Care J*. 2010;1(3):7–9. <https://doi.org/10.1055/s-0030-1267062>.
 18. Eagleman DM. Why public dissemination of science matters: a manifesto. *J Neurosci*. 2013;33(30):12147–9. <https://doi.org/10.1523/JNEUROSCI.2556-13.2013>.
 19. Euser AM, Zoccali C, Jager KJ, Dekker FW. Cohort studies: prospective versus retrospective. *Nephron Clin Pract*. 2009;113(3):c214–7. <https://doi.org/10.1159/000235241>.
 20. Eysenbach G. Can tweets predict citations? Metrics of social impact based on twitter and correlation with traditional metrics of scientific impact. *J Med Internet Res*. 2011;13(4):e123. <https://doi.org/10.2196/jmir.2041>.
 21. Farrugia P, Petrisor BA, Farrokhyar F, Bhandari M. Practical tips for surgical research: research questions, hypotheses and objectives. *Can J Surg*. 2009;53(4):278–81. <https://doi.org/10.1503/cjs.036311>.
 22. Freedman KB, Back S, Bernstein J. Sample size and statistical power of randomised, controlled trials in orthopaedics. *J Bone Joint Surg Br*. 2001;83(3):397–402. [papers3://publication/uuid/611B66F4-34E7-4087-878C-527960117276](https://pubmed.ncbi.nlm.nih.gov/117276/).
 23. Gofton WT, Solomon M, Gofton T, et al. What do reported learning curves mean for orthopaedic surgeons? *Instr Course Lect*. 2016;65:633–43.
 24. Haynes R. Forming research questions. In: Haynes R, Sackett D, Guyatt G, Tugwell P, editors. *Clinical epidemiology: how to do clinical practice research*. Philadelphia: Lippincott Williams & Watkins; 2006. p. 3–14.
 25. Helmy A, Timofeev I, Santarius T, Hutchinson P. What constitutes clinical equipoise. *Br J Neurosurg*. 2009;23(5):564–5. <https://doi.org/10.1080/02688690903029760>.
 26. Hewitt C. Adequacy and reporting of allocation concealment: review of recent trials published in four general medical journals. *BMJ*. 2005;330(7499):1057–8. <https://doi.org/10.1136/bmj.38413.576713.AE>.
 27. Higgins J, Green S. *Cochrane handbook for systematic reviews of interventions* version 5.1.0 [updated March 2011]. *Cochrane Collab*. 2011;0(March):10–11. <http://www.cochrane.org/training/cochrane-handbook>.
 28. Hip Fracture Accelerated Surgical Treatment and Care Track (HIP ATTACK) Investigators. Accelerated care versus standard care among patients with hip fracture: the HIP ATTACK pilot trial. *CMAJ*. 2014;186(1):E52–60. <https://doi.org/10.1503/cmaj.130901>.
 29. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. *Designing clinical research*, vol. 78. Philadelphia: Lippincott Williams & Wilkins; 2007.
 30. Karanicolas PJ, Bhandari M, Taromi B, et al. Blinding of outcomes in trials of orthopaedic trauma: an opportunity to enhance the validity of clinical trials. *J Bone Joint Surg Am*. 2008;90(5):1026–33. <https://doi.org/10.2106/JBJS.G.00963>.
 31. Karanicolas PJ, Farrokhyar F, Bhandari M. Blinding: who, what, when, why, how? *Can J Surg*. 2010;53(5):345–8. <https://doi.org/10.1007/s10269-005-0145-9>.
 32. Keating JF, Grant A, Masson M, Scott NW, Forbes JF. Displaced intracapsular hip fractures in fit, older people: a randomised comparison of reduction and fixation, bipolar hemiarthroplasty and total hip arthroplasty. *Health Technol Assess*. 2005;9(41):iii–iv, ix–x, 1–65.
 33. Kennedy A, Grant A. Subversion of allocation in a randomised controlled trial. *Control Clin Trials*. 1997;18(Suppl 3):S77–88.
 34. Leon AC, Davis LL, Kraemer HC. The role and interpretation of pilot studies in clinical research. *J Psychiatr Res*. 2011;45(5):626–9. <https://doi.org/10.1016/j.jpsychires.2010.10.008>.
 35. Lochner HV, Bhandari M, Tornetta P. Type-II error rates (beta errors) of randomized trials in orthopaedic trauma. *J Bone Joint Surg Am*. 2001;83(11):1650–5. <https://doi.org/10.2106/00004623-200111000-00005>.
 36. Louw A, Diener I, Fernández-de-las-Peñas C, Puentedura EJ. Sham surgery in orthopedics: a systematic review of the literature. *Pain Med*. 2017;18(4):736–50. <https://doi.org/10.1093/pm/pnw164>.
 37. Makel MC, Plucker JA. Facts are more important than novelty: replication in the education sciences. *Educ Res*. 2014;43(6):304–16. <https://doi.org/10.3102/0013189x14545513>.
 38. Malavolta EA, Demange MK, Gobbi RG, Imamura M, Fregni F. Randomized controlled clinical trials in orthopedics: difficulties and limitations. *Rev Bras Ortop*. 2011;46(4):452–9. [https://doi.org/10.1016/s2255-4971\(15\)30261-5](https://doi.org/10.1016/s2255-4971(15)30261-5).
 39. Mandal J, Acharaya S, Parija SC. Ethics in human research. *Trop Parasitol*. 2011;1(1):2–3.
 40. Marchant GE, Scheckel K, Campos-outcalt D. Contrasting medical and legal standards of evidence: a precision medicine case study. *J Law Med Ethics*. 2016;44(1):194–204. <https://doi.org/10.1177/1073110516644210>.
 41. Mehta S, Myers T, Lonner J, Huffma R, Sennett BJ. The ethics of sham surgery in clinical orthopaedic research. *J Bone Joint Surg Am*. 2007;89(7):1650–3.
 42. Misra S. Randomized double blind placebo control studies, the “gold standard” in intervention based studies. *Indian J Sex Transm Dis AIDS*. 2012;33(2):131–4. <https://doi.org/10.4103/0253-7184.102130>.
 43. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Int J Surg*. 2012;10(1):28–55. <https://doi.org/10.1016/j.ijssu.2011.10.001>.
 44. Mundi R, Chaudhry H, Mundi S, Godin K, Bhandari M. Design and execution of clinical trials in orthopaedic surgery. *Bone Joint Res*. 2014;3(5):161–8. <https://doi.org/10.1302/2046-3758.35.2000280>.

45. Nissen T, Wynn R. The history of the case report: a selective review. *JRSM Open*. 2014;5(4):5. <https://doi.org/10.1177/2054270414523410>.
46. Petrisor B, Jeray K, Schemitsch E, et al. Fluid lavage in patients with open fracture wounds (FLOW): an international survey of 984 surgeons. *BMC Musculoskelet Disord*. 2008;9(1):7. <https://doi.org/10.1186/1471-2474-9-7>.
47. Raynauld J-P, Buckland-Wright C, Ward R, et al. Safety and efficacy of long-term intraarticular steroid injections in osteoarthritis of the knee: a randomized, double-blind, placebo-controlled trial. *Arthritis Rheum*. 2003;48(2):370–7. <https://doi.org/10.1002/art.10777>.
48. Richardson WS, Wilson MC, Nishikawa J, Hayward RS. The well-built clinical question: a key to evidence-based decisions. *ACP J Club*. 1995;123(3):A12–3. <https://doi.org/10.7326/ACPJC-1995-123-3-A12>.
49. Röhrig B, du Prel J-B, Wachtlin D, Kwicien R, Bletner M. Sample size calculation in clinical trials. *Dtsch Arztebl Int*. 2010;107(31–32):552–6. <https://doi.org/10.3238/arztebl.2010.0552>.
50. Roman H, Marpeau L, Hulsey TC. Surgeons' experience and interaction effect in randomized controlled trials regarding new surgical procedures. *Am J Obstet Gynecol*. 2008;199(2):108.e1–6. <https://doi.org/10.1016/j.ajog.2008.03.002>.
51. Rothschild PC. Social media use in sports and entertainment venues. *Int J Event Festival Manag*. 2011;2(2):139–50. <https://doi.org/10.1108/17582951111136568>.
52. Schulz KF, Chalmers I, Hayes RJ, Altman DG. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA*. 1995;273(5):408–12. <https://doi.org/10.1001/jama.273.5.408>.
53. Schwitzer G. How do US journalists cover treatments, tests, products, and procedures? An evaluation of 500 stories. *PLoS Med*. 2008;5(5):e95. <https://doi.org/10.1371/journal.pmed.0050095>.
54. Selvaraj S, Borkar DS, Prasad V. Media coverage of medical journals: do the best articles make the news? *PLoS One*. 2014;9(1):e85355. <https://doi.org/10.1371/journal.pone.0085355>.
55. Smith OM, McDonald E, Zytaruk N, et al. Enhancing the informed consent process for critical care research: strategies from a thromboprophylaxis trial. *Intensive Crit Care Nurs*. 2013;29(6):300–9. <https://doi.org/10.1016/j.iccn.2013.04.006>.
56. Speich B. Blinding in surgical randomized clinical trials in 2015. *Ann Surg*. 2017;266(1):21–2. <https://doi.org/10.1097/SLA.0000000000002242>.
57. Sprague S, Quigley L, Bhandari M. Survey design in orthopaedic surgery: getting surgeons to respond. *J Bone Joint Surg Am*. 2009;91(Suppl 3):27–34. <https://doi.org/10.2106/jbjs.h.01574>.
58. Stieglitz S, Dang-Xuan L. Social media and political communication: a social media analytics framework. *Soc Netw Anal Min*. 2013;3(4):1277–91. <https://doi.org/10.1007/s13278-012-0079-3>.
59. Suresh K. An overview of randomization techniques: an unbiased assessment of outcome in clinical research. *J Hum Reprod Sci*. 2011;4(1):8. <https://doi.org/10.4103/0974-1208.82352>.
60. Thabane L, Thomas T, Ye C, Paul J. Posing the research question: not so simple. *Can J Anesth*. 2009;56(1):71–9. <https://doi.org/10.1007/s12630-008-9007-4>.
61. The FLOW Investigators. A trial of wound irrigation in the initial management of open fracture wounds. *N Engl J Med*. 2015;373:2629–41.
62. The FLOW Investigators. Fluid lavage of open wounds (FLOW): design and rationale for a large, multicenter collaborative 2 × 3 factorial trial of irrigating pressures and solutions in patients with open fractures. *BMC Musculoskelet Disord*. 2011;11:85.
63. Thoma A, Farrokhyar F, Mcknight L, Bhandari M. How to optimize patient recruitment. *Can J Surg*. 2010;53(3):205–10. <https://doi.org/10.1503/cjs.018012>.
64. Thoma A, McKnight L, McKay P, Haines T. Forming the research question. *Clin Plast Surg*. 2008;35(2):189–93. <https://doi.org/10.1016/j.cps.2007.10.009>.
65. Vavken P. Rationale for and methods of superiority, noninferiority, or equivalence designs in orthopaedic, controlled trials. *Clin Orthop Relat Res*. 2011;469(9):2645–53. <https://doi.org/10.1007/s11999-011-1773-6>.
66. Viswanath K, Blake KD, Meissner HI, et al. Occupational practices and the making of health news: a national survey of U.S. health and medical science journalists. *J Health Commun*. 2008;13(8):759–77. <https://doi.org/10.1080/10810730802487430>.
67. WMA. Declaration of Helsinki. Vol. 353. 1974. <https://doi.org/10.2471/BLT.08.057737>.



Level 1 Evidence: Long-Term Clinical Results

30

Daisuke Araki and Ryosuke Kuroda

30.1 Manuscript

Evidence-based medicine (EBM) is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients [8]. Since the Canadian Task Force on the periodic health examination originally described the level of evidence in 1979 [12], the level of evidence system is an evidence-based medicine tool that applies a hierarchical rating to a study's strength of evidence based on its study design. Nowadays, this system has been quite familiar, as its use has become widespread in medicine.

Regarding the evolution of EBM in orthopedics, the Journal of Bone and Joint Surgery, in recognition of the need to integrate clinical expertise with the best available systematic research, introduced "Evidence-Based Orthopaedics" in 2000 [10]. In the introduction to this new section, randomized clinical trials (RCT) would form the main contribution, because they are believed to provide the highest-quality evidence and therefore, when available, should influence clinical decision-making. Among the papers to appear in this section were a series of four levels of evidence for primary research question, covering prognosis, surgical therapies, diagnostic tests, and economic analyses [2–4]. These articles were

meant to teach orthopedic surgeons the manner in which evidence can be evaluated and applied in their practice. Among them, studies categorized as level 1 include RCTs and systematic reviews of level 1 RCTs in therapeutic studies, prospective studies and systematic reviews of level 1 studies in prognostic studies, testing of previously developed diagnostic criteria in series of consecutive patients and systematic reviews of level 1 studies in diagnostic studies, and clinically sensible costs and values obtained from many studies and systematic reviews of level 1 studies in economic and decision analyses [11].

Data derived from RCTs is considered to be the highest level of evidence, mainly because randomization is the best way to balance known and the only way to balance unknown prognostic factors within both treatment and control groups in a therapeutic study. On the other hand, it is also important to recognize that not all clinical questions can be answered with an RCT. While randomization in RCTs can be stratified based on prognostic factors, in some cases, it would be unethical to actively randomize patients to certain types of prognostic or risk factors. Prognostic factors of a disease or intervention can be assessed with a long-term follow-up study design, which then provides the highest level of evidence without being an RCT. In addition, there are other situations where an RCT may not be feasible, for example, when the sample size required is too large or the follow-up requires many years [7]. In

D. Araki (✉) · R. Kuroda
Department of Orthopedic Surgery, Kobe University
Graduate School of Medicine, Kobe, Japan
e-mail: isuke@pop21.odn.ne.jp

this context, we will focus on level 1 studies of the long-term clinical results.

Clinical Vignette 1

Study title (e.g.): 20 years' clinical follow-up and osteoarthritic change after anterior cruciate ligament (ACL) reconstruction. The title should include that the topic is important, relevant, and innovative. It is a summary of the abstract and should be short, descriptive, and interesting.

First, it is unethical to design the use of placebo and randomization regarding this topic. Patients assigned to the placebo arm of a clinical trial must be made to believe they are receiving a working treatment, even though they are not, for the placebo effect to play a role at all. However, a more serious issue with the use of placebo remains the possibility that participants are harmed by receiving a placebo instead of an active treatment [6]. If the patients are divided into two groups, (1) arthroscopic inspection and ACL reconstruction and (2) arthroscopic inspection only, the placebo may expose the patients to higher levels of pain and aggravate their condition.

Second, concepts of study methodology are crucial to consider when placing a study into the levels of evidence. There are some that advocate dividing the hierarchy levels into sub-levels based, in part, on study methodology. Others suggest that poor methodology will take a study down a level [1]. Therefore, in starting clinical research, researchers first create a "clinical research protocol."

Clinical Vignette 2

Is the patient cohort adequately reported, e.g., age, sex distribution, concomitant injuries, surgery, and sufficient power?

30.1.1 Study Design

The research protocols include (1) principal investigator who has the responsibility of the research, (2) clinical hypothesis, (3) research design, (4) assessment of progress status, (5)

evaluation of adverse event/effectiveness monitoring, and (6) statistical analysis. In particular, the research design should fully consider the ethical aspects of patient care, as well as the scientific aspects to accomplish the research and to properly verify the clinical hypothesis. Investigators formulated the study question before the first patient was enrolled. Next, identify the optimal number of cases. If the optimal number of cases is not determined, it may fail to obtain significant results (i.e., underpowered).

Clinical Vignette 3

Has informed consent been obtained? Is the study approved by the institutional review board or ethical committee?

30.1.2 Informed Consent

One of the most important ethical constructs of clinical ethics, informed consent, is an essential condition both for therapy and research. Information that makes consent valid is generally thought to include the understanding of the risks and benefits of the treatment(s) that patients may receive; understanding of the procedures that the participant may undergo, including, in the case of RCTs, blinding and randomization; understanding that participation in research is voluntary; and, finally, understanding the purpose of the research [6].

Clinical Vignette 4

Who performed the treatment? Who evaluated the patient outcome? Who performed the statistical analysis?

30.1.3 Blinding

Blinding is most often performed in prospective studies. Personnel who evaluate the outcomes of interest may have a belief or suspicion of which treatment offers the best outcome. They may interpret marginal results in a way that favors their presupposition if they are privy to the treatment administered. It is important to understand who is performing the data collection. Therefore,

those who are evaluating the results should be blinded to the treatment. Another way that bias can enter into the unblinded assessment process is through differential encouragement during a performance test [5].

Clinical Vignette 5

How long should be the follow-up period?

30.1.4 Follow-Up Rate and Period

The number lost to follow-up is very important to know as clearly this can affect the estimate of treatment effect. In general, the validity of a study may be threatened if more than 20% of patients are lost to follow up. Therefore, 80% follow-up of enrolled patients is necessary. Calculations of results should include a worst-case scenario; that is, those that are lost to follow-up are considered to have the worst outcome in the treatment group, and those that are lost to follow-up in the control group have the best outcome. If there is still a treatment effect seen between the groups, then this makes a more compelling argument for the treatment effect observed being a valid estimate of the truth. When conducting a clinical trial, researchers attempt to minimize data loss. Careful planning of research protocols, including comprehensive initial data collection, identification of locators, flexible scheduling, systematic subject tracking, monitoring subject loss, and systematically approaching problem cases, can ensure high follow-up rates [9].

Regarding follow-up period, precise definition of “long-term” has not been determined yet. However, according to the instructions of high-quality journal in orthopedics, the observation period in clinical research is specified to be minimum 2 years. Investigators should design the follow-up period along these instructions.

Fact Box

- The careful planning of research protocols are the key points to accomplish the high-quality clinical follow-up results.

- Comprehensive initial data collection, flexible scheduling, identification of locators, data blinding, systematic subject tracking, monitoring subject loss (follow-up rate $\geq 80\%$), follow-up period (≥ 2 years), and systematically approaching problem cases are mandatory for the level 1 study.
- It is also essential to analyze the data with adequate statistics and sample size.

References

1. Atkins D, Best D, Briss PA, et al. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328(7454):1490.
2. Bhandari M, Guyatt GH, Swiontkowski MF. User's guide to the orthopaedic literature: how to use an article about a surgical therapy. *J Bone Joint Surg Am*. 2001;83-A(6):916–26.
3. Bhandari M, Guyatt GH, Swiontkowski MF. User's guide to the orthopaedic literature: how to use an article about prognosis. *J Bone Joint Surg Am*. 2001;83-A(10):1555–64.
4. Bhandari M, Morrow F, Kulkarni AV, Tornetta P 3rd. Meta-analyses in orthopaedic surgery. A systematic review of their methodologies. *J Bone Joint Surg Am*. 2001;83-A(1):15–24.
5. Dettori J. Class or level of evidence: epidemiologic basis. *Evid Based Spine Care J*. 2012;3(3):9–12.
6. Nardini C. The ethics of clinical trials. *Ecancermedalscience*. 2014;8:387.
7. Poolman RW, Petrisor BA, Marti RK, Kerkhoffs GM, Zlowodzki M, Bhandari M. Misconceptions about practicing evidence-based orthopedic surgery. *Acta Orthop*. 2007;78(1):2–11.
8. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *BMJ*. 1996;312(7023):71–2.
9. Woolard RH, Carty K, Wirtz P, et al. Research fundamentals: follow-up of subjects in clinical trials: addressing subject attrition. *Acad Emerg Med*. 2004;11(8):859–66.
10. Wright JG, Swiontkowski MF. Introducing a new journal section: evidence-based orthopaedics. *J Bone Joint Surg Am*. 2000;82(6):759–60.
11. Wright JG, Swiontkowski MF, Heckman JD. Introducing levels of evidence to the journal. *J Bone Joint Surg Am*. 2003;85-A(1):1–3.
12. The periodic health examination. Canadian Task Force on the Periodic Health Examination. *Can Med Assoc J*. 1979;121(9):1193–254.

Level 2 Evidence: Prospective Cohort Study

31

Naomi Roselaar, Niv Marom, and Robert G. Marx

31.1 Introduction

31.1.1 Levels of Evidence

Levels of evidence are used in evidence-based medicine to categorize clinical research [4]. Level 1 evidence is considered the most rigorous and is assigned to randomized controlled trials with proper blinding, appropriate randomization, and high follow-up rates. Level 2 evidence is considered slightly less rigorous and is assigned to prospective cohort studies. This is considered more rigorous than level 3 evidence, retrospective cohort studies, or level 4 evidence, case series and case studies. The hierarchy of levels of evidence is based on the risk of bias and systematic error in each type of study [4]. However, studies comparing treatment effects among randomized and observational studies showed neither consistent nor systematic overestimation of treatment effects in nonrandomized studies [1, 5].

Levels of evidence are a relatively new standard of classification in orthopedic research.

The *Journal of Bone and Joint Surgery* began assigning all articles a level of evidence in 2003 [17]. The *Journal of Orthopedic Trauma* intro-

duced the assignment of levels of evidence to all clinical articles in 2012 [15].

In addition to aiding classification, levels of evidence help peer-reviewed orthopedics journals maintain high standards of research quality. Improving quality was the primary reason for assigning levels of evidence in the *Journal of Hand Surgery* beginning in November 2005 [8]. However, the quality of studies published in the *Journal of Bone and Joint Surgery* (American) increased over time even before the introduction of levels of evidence [6]. In 2009, a retrospective assessment found that between 1975 and 2005, the percentage of level 1 studies published in the *Journal of Bone and Joint Surgery* (American) increased from 4 to 21% [6].

31.1.2 Prospective Cohort Study

Level 2 prospective cohort studies are useful for assessing outcomes in patients with different treatments or characteristics [7]. As such, level 2 prospective cohort studies have either a control group or second, distinct treatment group for comparison against the primary treatment group. Prospective cohort studies can be therapeutic or prognostic [9]. Prospective cohort studies require the research hypothesis and question to be determined prior to recruitment and enrollment of participants [16]. With the predetermined research question, investigators obtain data about the

N. Roselaar · N. Marom · R. G. Marx (✉)
Hospital for Special Surgery, New York, NY, USA
e-mail: roselaarn@hss.edu; maromn@hss.edu;
marxr@hss.edu

patient and intervention from the beginning of the study. Outcomes that occur over the course of the study period are assessed [3].

Fact Box 31.1

In level 2 prospective cohort studies, the research question and hypothesis are established prior to recruitment and enrollment of subjects.

31.2 Benefits and Limitations

31.2.1 Benefits

Prospective cohort studies are observational, which makes them less difficult to execute than randomized controlled trials. They also tend to be less disruptive to patient flow within a clinic. Logistically, coordinating an observational study requires less interference with standard clinical practices than does coordinating a study with proper blinding and randomization [3].

When designing prospective cohort studies, investigators determine data collection methods specific to their research question prior to enrolling patients. This high level of specification in the method of data collection gives investigators greater control over the data collected compared to retrospective studies. It also ensures specificity in the data collected [16]. The likelihood of reporting relevant risk factors and outcomes is higher in prospective cohort studies [12]. With retrospectively collected data, the research question is limited by existing data [16].

Prospective cohort studies minimize recall bias because data is collected longitudinally, in real time. In prospective studies, the data collection occurs as the outcome of interest progresses [12], whereas in retrospective case-control studies, case (those with a disease) and control (healthy) participants are susceptible to providing information with varying degrees of accuracy on their exposure to the disease or outcome of interest [13].

Prospective cohort studies also allow investigators to examine multiple outcomes simultaneously [16].

31.2.2 Limitations

Prospective studies have the potential for decreased internal validity because of susceptibility to selection bias. Unlike randomized controlled trials, prospective cohort studies may use patient or physician preference when determining treatment allocation. When assessing the differences in outcomes among treatment and control groups, factors that went into deciding which intervention each patient received must also be considered to mitigate selection bias [3].

Longitudinal studies, such as prospective cohort studies, may have higher proportions of subjects lost to follow-up. This contributes to bias if the reasons for loss of follow-up are related to the outcome [12]. A low follow-up rate that impacts study quality may decrease the level of evidence of the study [9]. For studies concerning rare diseases or diseases with long latency, designing a prospective cohort study may be impractical and/or unfeasible [16].

31.3 Controlled Prospective Studies

31.3.1 Controlled Prospective Cohort Study

In a controlled prospective cohort study, the treatment group is compared to a control group or nontreatment group.

Clinical Vignette 1

“Telephone-Based Intervention to Improve Rehabilitation Engagement After Spinal Stenosis Surgery” (JBJS, 2018).

Investigators from the Departments of Orthopaedic Surgery, Physical Medicine, and Rehabilitation at the Johns Hopkins University

School of Medicine designed an uncontrolled prospective cohort study to assess rehabilitation engagement following spinal stenosis surgery [14]. Their goal was to compare the effectiveness of the “usual care” to a new telephone-based counseling system on rehabilitation engagement. In this study, “usual care” included physical therapy following surgical treatment and a follow-up appointment with a physical exam and assessment of radiographic imaging. The 60 patients that were prospectively enrolled into “usual care” made up the control group. All subjects in both the control and intervention groups were enrolled before undergoing surgery. The intervention group comprised 65 patients who received 3 health behavior change counseling phone calls (one preoperatively and two postoperatively) in addition to physical therapy and a follow-up assessment. According to the study, the health behavior change counseling phone calls included “motivational interviewing strategies that elicit and strengthen motivation for change.” For patients receiving this telephone-based intervention, the investigators hypothesized better health outcomes related to rehabilitation. As expected, the intervention group showed better pain, disability, and physical health outcomes at 12 months, likely due to physical therapy engagement and attendance. During the second and third years that all subjects were followed, differences in outcomes between the control and intervention groups decreased [14].

This level 2 prospective cohort study was an appropriate method for assessing patient engagement. Prior to beginning the study, investigators determined that subjects would be assigned to either the control or intervention group. Once enrolled, all patients completed outcome questionnaires, which allowed investigators to follow the patient outcomes over time. This was not a randomized cohort study, because the group assignments were decided by enrollment date. Importantly, in this controlled study, the investigators determined that the control group was provided sufficient care without receiving the intervention. The care received by the control group was standard of care.

Orthopedic surgeons from the Sungkyunkwan University School of Medicine in Seoul, Korea,

Clinical Vignette 2

Treatment of Biceps Tendon Lesions in the Setting of Rotator Cuff Tears – Prospective Cohort Study of Tenotomy Versus Tenodesis (AJSM 2010).

compared postsurgical functional outcomes in patients presenting with rotator cuff tears and biceps tendon lesions [11]. Of the 90 subjects included in the study, 45 underwent rotator cuff repair with biceps tenotomy, and 45 were treated with rotator cuff repair and biceps tenodesis. Subjects in both groups received the same postoperative immobilization and rehabilitation. To assess functional outcomes and satisfaction, investigators administered physical exams and questionnaires to both intervention groups at follow-up appointments. A significant statistical difference in the number of patients with Popeye deformity was found between those in the tenodesis (9.3%) and tenotomy (26.8%) groups. Functional outcomes, clinical scores, and total surgical times were similar for the two groups.

In this prospective cohort study, the outcomes of two different surgical interventions (tenotomy and tenodesis) for the same diagnosis (biceps tendon lesions) were compared. Like the study of telephone-based rehabilitation engagement, patient groups were determined according to each subject’s enrollment date [11].

Fact Box 31.2

Controlled prospective cohort studies involve either a comparison between a treatment group and a control group or between two different treatment groups.

31.4 Prognostic Prospective Cohort Study

Prognostic studies assess the effect of a patient characteristic on the outcome of a disease [9]. Patient characteristics can be behavioral, such as level of athletic training or smoking status; physi-

cal, as in classes of obesity; or based on a genetic trait. Prognostic studies answer the question “What is the effect of a patient characteristic on the natural history of the condition?” [9].

Clinical Vignette 3

No Effect of Generalized Joint Hypermobility on Injury Risk in Elite Female Soccer Players – A Prospective Cohort Study (AJSM, 2017).

In a 2016 study from the Netherlands, the effect of generalized joint hypermobility (GJH) on injury rate was investigated in elite female soccer players [2]. In this prognostic study, hypermobility due to GJH is the patient characteristic, while injury rate measured in player injuries per hours of soccer is the outcome of interest. Athletes were classified as hypermobile or non-hypermobile after screening using the Beighton score. With a Beighton score ≥ 4 , 20 of the 114 athletes enrolled in the study were classified as hypermobile. The number of injuries accrued by standardized injury registration forms was assessed, and non-musculoskeletal injuries such as concussions as well as non-soccer-related injuries were excluded. At the end of one soccer season, investigators concluded that GJH was not a risk factor for injuries in elite female soccer players. These results were maintained when the Beighton score threshold for hypermobility was ≥ 3 , ≥ 4 , and ≥ 5 [2].

In this case the comparison groups were determined by dividing a spectrum (Beighton score) to produce two groups for comparison. In addition to GJH, classes of obesity are another example of spectrum-based comparison groups. Comparison groups for prognostic studies also include dichotomous patient characteristics, such as being hemophilic vs not being hemophilic [10].

Prognostic prospective cohort studies can be classified as level one if they are inception cohort studies [9]. Inception studies require all subjects to be enrolled at the same point in their disease. This was not the case in the above example. In the study from the Netherlands, injuries were the

outcome of interest. Pre-existing injuries sustained by participating athletes were disregarded in the injury rate calculation. The presence of varying levels of exposure to injury at the beginning of the study indicates level two evidence [9].

Take-Home Message

- Uncontrolled or controlled, therapeutic or not, prospective cohort studies evaluate patient outcomes in real time over the course of the study.
- Classified as level 2 evidence, a prospective therapeutic cohort study is the most rigorous research method aside from a randomized clinical trial.

References

1. Benson K, Hartz A. A comparison of observational studies and randomized, controlled trials. *N Engl J Med.* 2000;342(25):1878–86.
2. Blokland D, Thijs KM, Backx FJG, Goedhart EA, Huisstede BMA. No effect of generalized joint hypermobility on injury risk in elite female soccer players: a prospective cohort study. *Am J Sports Med.* 2017;45(2):286–93. <https://doi.org/10.1177/0363546516676051>.
3. Bryant DM, Willits K, Hanson BP. Principles of designing a cohort study in orthopaedics. *J Bone Joint Surg Ser A.* 2009;91:10–4. <https://doi.org/10.2106/JBJS.H.01597>.
4. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg.* 2011;128(1):305–10. <https://doi.org/10.1097/PRS.0b013e318219c171>.
5. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med.* 2000;342(25):1887–92. <https://doi.org/10.1056/NEJM200006223422507>.
6. Hanzlik S, Mahabir RC, Baynosa RC, Khiabani KT. Levels of evidence in research published in *The Journal of Bone and Joint Surgery (American Volume)* over the last thirty years. *J Bone Joint Surg.* 2009;91(2):425–8. <https://doi.org/10.2106/JBJS.H.00108>.
7. Hennekens C, Buring J. *Epidemiology in medicine.* 1st ed. Boston: Little, Brown; 1987.
8. Hentz R, Meals R, Stern P, Manske P. Levels of evidence and the journal of hand surgery. *J Hand Surg Am.* 2015;30(5):891–2.
9. JBJS Inc. Journals level of evidence. *The Journal of Bone and Joint Surgery.* <https://journals.lww.com/jbjsjournal/Pages/Journals-Level-of-Evidence.aspx>. Published 2015.

10. Kapadia BH, Boylan MR, Elmallah RK, Krebs VE, Paulino CB, Mont MA. Does hemophilia increase the risk of postoperative blood transfusion after lower extremity total joint arthroplasty? *J Arthroplast.* 2016;31:1578–82. <https://doi.org/10.1016/j.arth.2016.01.012>.
11. Koh KH, Ahn JH, Kim SM, Yoo JC. Treatment of biceps tendon lesions in the setting of rotator cuff tears: prospective cohort study of tenotomy versus tenodesis. *Am J Sports Med.* 2010;38(8):1584–90. <https://doi.org/10.1177/0363546510364053>.
12. Sedgwick P. Prospective cohort studies: advantages and disadvantages. *BMJ.* 2013;347:f6726. <https://doi.org/10.1136/bmj.f6726>.
13. Sedgwick P. What is recall bias? *BMJ.* 2012;344:e3519.
14. Skolasky RL, Maggard AM, Wegener ST, Riley Iii LH. Telephone-based intervention to improve rehabilitation engagement after spinal stenosis surgery: a prospective lagged controlled trial. *J Bone Joint Surg.* 2018;100:21–30. <https://doi.org/10.2106/JBJS.17.00418>.
15. Slobogean G, Bhandari M. Introducing levels of evidence to the journal of orthopaedic trauma: implementation and future directions. *J Orthop Traumatol.* 2012;26(3):127–8.
16. Song JW, Chung KC. Observational studies: cohort and case-control studies. *Plast Reconstr Surg.* 2010;126(6):2234–42. <https://doi.org/10.1097/PRS.0b013e3181f44abc>.
17. Wright JG, Swiontkowski MF, Heckman JD. Introducing levels of evidence to the journal. *J Bone Joint Surg.* 2003;85(1):1–3.



Level III Evidence: A Case-Control Study

32

Andrew D. Lynch, Adam J. Popchak,
and James J. Irrgang

32.1 Introduction

Observational research, including case-control studies, are important study designs that contribute to both general knowledge and hypothesis generation [1]. A case-control study departs from the standard process of enrollment and prospective observation in randomized controlled trials and observational cohorts. In a case-control study, participants who are known to have the condition or outcome of interest are chosen as cases. Potential contributors to the etiology of the disease are sought by looking backward in the history of the participant to find out if something in their history may have caused the disease or condition of interest [2, 6, 9, 12]. This has led some to consider a case-control study as “research in reverse” [11]. The precipitating factor that may have contributed to the development of the outcome is called an exposure (also independent variable, determinant, or predictor) [12]. The exposure may be a biologic exposure, some characteristic of the individual, some behavioral event, or an intervention [11].

The distinct advantage of a case-control study over a simple case series is the inclusion of a group of control participants who are largely sim-

ilar to the case participants. Control participants are chosen because they do not have the condition of interest but are chosen from the same population as the cases. Because they come from the same population, it is assumed that differences in the exposure between the cases and controls are associated with the development of the condition or outcome of interest [5]. Case-control studies are especially important when the researcher cannot randomize a patient to a particular exposure based on their clinical judgment (e.g., randomizing a patient to non-operative treatment if the researcher does not believe this is best for the patient) [8].

It is important to differentiate a case-control study from a prospective cohort study because both make the argument that an exposure causes an outcome [5]. In a case-control study, participants are chosen based on whether or not they have the outcome of interest, and their history is reviewed retrospectively to determine their history of exposure. In a cohort study, participants are chosen based on their exposure status and followed prospectively to determine whether they develop the outcome of interest. A review of manuscripts purported to report the results of “case-control studies” found that about one-third of medical/surgical “case-control” studies were incorrectly labeled and 97% of rehabilitation reports were mislabeled [6]. The most frequent mislabeling was attributed to a cross-sectional study design, with intervention studies,

A. D. Lynch (✉) · A. J. Popchak · J. J. Irrgang
Departments of Physical Therapy and Orthopaedic
Surgery, University of Pittsburgh,
Pittsburgh, PA, USA
e-mail: adl45@pitt.edu

measurement studies, prognostic studies, and cohort studies all being incorrectly labeled as case-control studies.

Fact Box 32.1: Key Definitions

Case: a research participant who is known to have the condition of interest.

Control: a research participant who does not have the condition of interest but is otherwise similar to the case participants (e.g., selected from the same population as the cases).

Exposure: some prior experience for both cases and controls that is hypothesized to lead to the development of the condition of interest. The primary hypothesis in a case-control study is that cases and controls will differ on their exposure history and that difference contributes to the condition of interest.

32.2 Advantages of Case-Control Studies

A case-control study is a preferable study design when considering outcomes that are either rare or that take a long time to develop, which would make a prospective study not feasible [8]. By selecting cases who have already developed the outcome of interest, the researcher knows that there will be a sample who has the outcome of interest. To prospectively study rare conditions, the researcher would have to recruit a large number of participants and spend resources to follow them to identify those who develop the condition of interest. To prospectively study conditions which take a long time to develop, the researcher would have to follow a cohort for a very long time. By choosing a study population who has already developed the outcome of interest, the researcher saves two precious resources—time and money [11].

Fact Box 32.2: Advantages and Disadvantages of Case-Control Studies

Advantages:

- Retrospective study design decreases the amount of time and money needed to complete a study to determine the role of an exposure in the etiology of a condition.
- Specific cohorts of individuals with a rare disease or outcome can be recruited to ensure a large enough sample size for analysis.

Disadvantages:

- Significant potential for bias to be introduced (see additional Fact Box 32.3).
- Does not benefit from thorough prospective data collection, thus data is subject to recall bias from participants and the quality of the medical record.
- Because the outcome is rare, it is usually not possible to measure the true incidence and prevalence of the outcome.
- There is significant risk of misclassification of the condition when expensive or invasive tests are not routinely used.

32.3 Disadvantages of Case-Control Studies

The retrospective nature of case-control studies also presents disadvantages, namely, through several sources of bias [7, 9, 10]. Bias occurs when something about the research design or conduct introduces error into the data set [12]. A case-control study does not benefit from standardized data collection methods early in the process, but instead relies on medical chart review and patient recall to provide information about the development of the condition. Medical records are summaries of the interaction between provider and patient and capture the details deemed relevant by

the provider. Medical records are not exhaustive records that are monitored for completeness in the same fashion that prospective study case-report forms are monitored. Therefore, there may be *information bias* associated with the reliance on the medical record as the source of data for a study—more information may be available for a patient who is having negative repercussions associated with treatment [11].

Additionally, anything that happens to the patient outside of the medical encounter will not be documented in the medical record. For this reason, participants are frequently asked to complete surveys about their experiences and exposures. The recall of the cases and controls may be biased differently—a case participant may have thought intensely about the scenarios that contributed to the development of the condition, while the control participant may not, leading the case subject to provide more detailed and accurate information or information that has been overanalyzed by the case participant [2, 9, 11, 12]. Additionally, survey data relies on the participants being willing to accurately report their experiences. In both cases and controls, *recall bias* could introduce potential error into the data set [7, 11].

There is also the potential for **selection bias** on the part of the researcher. The most simple design to execute involves cases and controls who have received care from a single surgeon or group practice. However, in selecting cases from a single surgeon or group practice, the researcher only has access to potential cases who chose to follow up with that surgeon or group. The researcher may not have access to a case who chose to go elsewhere for further care. In the case of a practice that specializes in the rare outcome of interest, it is unlikely that the practice will have performed the original care and may not have complete access to the medical record, leading to an incomplete data set. Cases and controls must be selected from a similar population and not in a way that introduces unwanted bias into the data set.

The case-control design clearly lacks the rigorous attention to detail in data collection of a

prospective cohort or randomized controlled trial, but this is balanced by the inexpensive and potentially rapid completion of the study. These should not be considered “fatal flaws” but should be factored into the analysis and interpretation of the data and results, as well as the overall discussion [12, 13]. These can be accounted for with good research practice, including the creation of detailed medical chart abstraction forms, structured design of retrospective surveys, adhering to strict inclusion and exclusion criteria for selecting cases and controls, and ensuring that data collectors are blinded to case vs. control status whenever possible [2, 11, 12].

Fact Box 32.3: Sources and Types of Bias

Information/observation bias occurs when information is gathered differently for the case participants and control participants. This may be due to differences in interview techniques, chart review methods, or survey design that introduces bias in support of the research hypothesis.

Selection bias occurs when case participants and control participants are chosen via different selection criteria. Specifically, case participants and control participants must be selected without consideration of their exposure history or any variable associated with the exposure.

Recall bias occurs when something particular to the cases or controls influences the recall such that more detailed information is gathered from one group.

32.4 Sample Selection

When selecting cases and controls, the researcher must select participants who are comparable in terms of both baseline risk of developing the outcome and the potential to have a complete data set related to that individual. It is important that they come from the same general population and have the same general characteristics.

32.4.1 Selecting Cases

When selecting cases, the researcher should attempt to identify a homogenous group. The condition of interest for those selected as cases should be diagnosed in a consistent fashion via an appropriate combination of diagnostic tests and clinical findings, including the gold standard as often as possible [11]. The diagnostic criteria for identifying a case should be established when planning the study, and all potential avenues for diagnosing the condition of interest should be considered. Misclassification of controls as cases (and vice versa) must be avoided when possible, accounted for in the design phase and sample size assessment [4], or handled with statistical sensitivity analyses (see Sect. 32.5) [3].

In orthopedics, cases are often selected from the practice of the researcher, but to identify a large enough sample to make some meaningful inferences about the condition, the researcher may need to seek out others to contribute cases to the sample. Clearly defined diagnostic criteria make this process more objective and the results more generalizable.

For chronic conditions, the researcher must decide whether to include only incident cases (newly or recently diagnosed) or whether prevalent cases (those who have been living with a condition for some time) may be included [2, 11]. For a condition such as osteoarthritis, there is a chronic, degenerative process that occurs. Including both an individual with significant osteoarthritis of the knee joint that has been worsening for years and an individual who has radiologic degeneration without symptoms can lead to substantially different reports of their exposure history.

32.4.2 Selecting Controls

As previously mentioned, controls should be selected to be similar to the cases in all respects except having the outcome of interest [11]. Ideally, the control subjects should also be representative of all individuals without the outcome

of interest in the population from which the cases are selected—for example, when selecting cases who develop a particular postoperative complication, the cases should be selected from all participants who had the same surgery but who did not develop the complication [7, 11]. This promotes the controls to have the same baseline risk as the cases. Any exclusionary criteria applied to cases should be applied equally to controls [7, 11, 12].

In a matched case-control study, the investigator matches the cases and controls for factors about which there is potential concern for confounding [2, 7, 12]. In group or frequency matching, the proportion of controls with a given characteristic or trait is identical to proportion of cases with that trait. In a matched pairs design, for each case selected, a control is selected who is similar in terms of variables of concern (e.g., age, sex, smoking status). These designs are more tightly controlled but also present issues with analysis and interpretation, especially when cases and controls are overmatched. Overmatching (matching on too many variables) eliminates potential exposures that may contribute to the outcome of interest [7]. As an example, matching cases and controls who have early-stage osteoarthritis for injury history related to the meniscus would eliminate the potential contribution of the meniscus injury as a contributor to development of osteoarthritis.

Because the case-control design is often used to study rare conditions, the sample of cases is often small. This is often not the case for the control subjects, so a matching of two or three control participants to each case participant may be used [7, 12]. This can increase the power of the study by including a more generalizable cohort of controls. This also improves the chances that the control group will have exposure similar to the case group. For studies in orthopedics specifically dealing with reinjury after primary surgery, it should be quite easy to identify a large group of patients with an index surgery who have not been reinjured. This will provide more variability in the control dataset and allow for more robust comparisons.

32.5 Statistical Analysis

A case-control study seeks to understand whether some exposure (disease, procedure, condition, or patient characteristic) has any effect on the probability of developing an outcome of interest [7, 9]. When reviewing the history of the cases and controls, the presence or absence of an exposure should be obtained from the medical record or via participant survey/interview. Then, each case (outcome positive) and control (outcome negative) can be classified as having been exposed (exposure positive) or not (exposure negative). This allows us to create a simple 2×2 table or *contingency table* to illustrate the difference between the exposed and unexposed (Fig. 32.1). We will discuss some simple statistics that can be used with a 2×2 table but recommend the researcher follow-up with a biostatistician for more complex, multivariable analysis. Additionally, methods for sensitivity analysis and issues related to sample size estimation are available but are beyond the scope of this chapter [3, 4].

Typically, contingency tables are used to calculate a *risk ratio* or *relative risk (RR)* to convey how exposure to a predictor puts someone at risk for developing the outcome of interest. This is done by dividing the population incidence of the condition of interest in those exposed [i.e., the proportion of exposed individuals who developed the condition (e.g., # of cases/# of exposed or $a/a + b$)] by the population incidence of the condition of interest in nonexposed individuals [the proportion of unexposed individuals who developed the condition (e.g., # of cases/# of nonexposed or $c/c + d$)]. However, the relative risk is not appropriate in a case-control study because the selection of cases and controls does not accurately reflect the true population incidence [11].

In the case-control study, an *odds ratio (OR)* can be used. Note that this is a less precise method

used in smaller samples where the true incidence of the condition of interest is rare (generally less than 5%) but still produces an estimate of risk [4, 11]. An odds ratio is determined with the following formula:

$$OR = \frac{\frac{a}{b}}{\frac{c}{d}} = \frac{ad}{bc}$$

Because the formula reduces down to a multiplication of opposite corners, it is sometimes referred to as a cross product. An odds ratio or relative risk greater than 1 would indicate that exposure increases the risk of developing the outcome of interest, while an odds ratio or relative risk less than 1 would indicate that exposure decreases the risk of developing the outcome of interest (i.e., the exposure is protective).

32.6 Reporting of Case: Control Studies

When reporting results of a case-control study, and even in the planning of a case-control study, researchers are recommended to consider the checklist from the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) consortium, a standardized guideline for the reporting of epidemiological studies in medicine [12, 13]. The STROBE statement has guidelines for report titles, abstracts, introduction, methods, results, and discussion sections.

32.7 Summary

Case-control studies should be used to retrospectively determine the role of an exposure in the etiology of an outcome or condition of interest

Fig. 32.1 Standard 2×2 table format

	Has the Condition of Interest	Does not have the Condition of Interest
Exposed	a	b
Unexposed	c	d

that is rare or takes a long time to develop. Because of the retrospective nature, case-control studies can be completed relative quickly and at a smaller cost than a prospective observational study. However, the retrospective nature may introduce multiple types of bias into the data set, and results must be considered in light of the limitations of the retrospective study design.

Participants are chosen based on whether they have the condition of interest (cases) or do not (controls) without regard to exposure. The exposure is ascertained via chart review, retrospective patient survey, or patient interview. An odds ratio is used to determine whether exposure increases risk (i.e., odds ratio greater than 1.0) or is protective (i.e., odds ratio less than 1.0). Despite the relative ease with which these studies can be completed, significant rigor should be put into place when designing these studies.

32.8 Useful Resources

STROBE Statement Web Site—<http://www.strobe-statement.org>

References

- Bhandari M, Morshed S, Tornetta P 3rd, Schemitsch EH. Design, conduct, and interpretation of nonrandomized orthopaedic studies: a practical approach. (All) evidence matters. *J Bone Joint Surg Am.* 2009;91(Suppl 3):1. <https://doi.org/10.2106/JBJS.H.01747>.
- Busse JW, Obremskey WT. Principles of designing an orthopaedic case-control study. *J Bone Joint Surg Am.* 2009;91(Suppl 3):15–20. <https://doi.org/10.2106/JBJS.H.01570>.
- Gilbert R, Martin RM, Donovan J, Lane JA, Hamdy F, Neal DE, Metcalfe C. Misclassification of outcome in case-control studies: methods for sensitivity analysis. *Stat Methods Med Res.* 2016;25:2377–93. <https://doi.org/10.1177/0962280214523192>.
- Joseph L, Belisle P. Bayesian sample size determination for case-control studies when exposure may be misclassified. *Am J Epidemiol.* 2013;178:1673–9. <https://doi.org/10.1093/aje/kwt181>.
- Mayo NE, Goldberg MS. When is a case-control study a case-control study? *J Rehabil Med.* 2009;41:217–22. <https://doi.org/10.2340/16501977-0341>.
- Mayo NE, Goldberg MS. When is a case-control study not a case-control study? *J Rehabil Med.* 2009;41:209–16. <https://doi.org/10.2340/16501977-0343>.
- Morshed S, Tornetta P 3rd, Bhandari M. Analysis of observational studies: a guide to understanding statistical methods. *J Bone Joint Surg Am.* 2009;91(Suppl 3):50–60. <https://doi.org/10.2106/JBJS.H.01577>.
- Nesvick CL, Thompson CJ, Boop FA, Klimo P Jr. Case-control studies in neurosurgery. *J Neurosurg.* 2014;121:285–96. <https://doi.org/10.3171/2014.5.JNS132329>.
- Portney LG, Watkins MP. Foundations of clinical research : applications to practice. 2nd ed. Upper Saddle River, NJ: Prentice Hall; 2000.
- Sackett DL. Bias in analytic research. *J Chron Dis.* 1979;32:51–63. [https://doi.org/10.1016/0021-9681\(79\)90012-2](https://doi.org/10.1016/0021-9681(79)90012-2).
- Schulz KF, Grimes DA. Case-control studies: research in reverse. *Lancet.* 2002;359:431–4. [https://doi.org/10.1016/S0140-6736\(02\)07605-5](https://doi.org/10.1016/S0140-6736(02)07605-5).
- Vandenbroucke JP, et al. Strengthening the reporting of observational studies in epidemiology (STROBE): explanation and elaboration. *PLoS Med.* 2007;4:e297. <https://doi.org/10.1371/journal.pmed.0040297>.
- von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, Initiative S. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med.* 2007;4:e296. <https://doi.org/10.1371/journal.pmed.0040296>.



Level 4 Evidence: Clinical Case Series

33

Mitchell I. Kennedy and Robert F. LaPrade

33.1 Introduction

There are many approaches to conducting clinical research for determining trends associated with variable conditions, ranging from the incidence, causality, and outcomes following treatment. Studies are assessed by the level of evidence presented in their findings and proceed as follows: controlled trials (Level I), cohort studies (Level II), case-control studies (Level III), case series (Level IV), and expert opinion (Level V). Two broad variations in clinical research include all clinical research studies, being classified as either analytic or descriptive [10]. Analytic studies are premised on confirming/disproving a hypothesis that aims to determine a causal relation between an exposure and the outcomes of a condition [10]. They can be designed as observational or controlled, resulting in either the direct selection or randomization of treatment, respectively, by the principal investigator, for the patients of the study [10]. Analytic

studies include case-control or cohort (observational) and randomized controlled trials (controlled). Case reports and case series are “descriptive,” observational studies that describe general disease characteristics associated with patients, places, and time [10]. The level of evidence is primarily distinguished by the effects that variable biases inflict upon the overall validity of the study’s findings.

Reporting Novel Strategies

- Aid in developing hypothesis for studies of greater level of evidence
- Exclusively report outcomes following a novel treatment procedure of a study population

Case series are unlike cohort and case-control studies in that they do not test a hypothesis or make use of a comparison group to determine the efficacy of a treatment. Rather, case series follow a group of patients over a period of time who have a similar diagnosis or are being treated with the same procedure [3, 10]. Case series are vital for analyzing unusual occurrences of a disease or designing a hypothesis for a potential prospective study [9]. This chapter is structured for defining the process of clinical research using a case series, primarily focusing on the process of design, reporting of outcomes, and the strengths and limitations of conducting research.

M. I. Kennedy
The Steadman Philippon Research Institute,
Vail, CO, USA
e-mail: mkennedy@sprivail.org

R. F. LaPrade (✉)
The Steadman Clinic, Vail, CO, USA

33.2 Design

Before any study may begin, it must be approved in accordance with the privacy law with approval from the respective Institutional Review Board (IRB). This law essentially protects any individual, living or deceased, of their protected health information and the confidentiality of data that is collected [12]. Additionally, the use of an individual's identifiable data may not be used without prior written approval [12]. A waiver may be granted for prior written approval required by the Privacy Rule if the following criteria are met: the project poses a minimal risk to the privacy of the individuals, there is an adequate plan to protect identifiers, the identifiers will be destroyed at the earliest opportunity, the project cannot be practically conducted without the specified protected health information, and the project could not be conducted without a waiver [12]. The approval by the respective local IRB usually comes following a thorough description of the approach and ultimately the goals for conducting the study. The key features of a case series that should be explicitly stated and focused around are a noteworthy collection of clinical occurrences that potentially display a combination of signs/symptoms and the subsequent novel treatment protocol that may infer causality [7].

Repeatability

- Highly descriptive techniques are necessary for comparison/analyses of future prospective studies
- Unambiguously stated inclusion/exclusion criteria for comparisons among differing institutions
- Explicitly stated indications for creating a consistent patient group

Case series are best utilized for reporting novel diagnostic/therapeutic strategies when the alternative option of delaying for comparative evidence is less likely [10]. Due to the lack of hypothesis or comparative groups, case series are unable to make causal inferences between a treatment method and its outcomes, but it can aid in

developing hypotheses for studies of a greater level of evidence [10]. Instead of typical clinical research studies comparing the effectiveness of relative procedures, case series exclusively report the outcomes following a novel treatment procedure for a specific study population [10]. Repeatability is very important regarding case series studies, because the incidence of novel procedures subsequent to unusual disease patterns is rarely standardized, and highly descriptive techniques are necessary for comparisons/analyses of future prospective studies. Unambiguously stated inclusion/exclusion criteria are also necessary for physicians of differing institutions to make comparisons with their own patient populations; defined and short inclusion periods are highly recommended to reduce the incidence of known and unknown changes that may occur more frequently in patients seen over varying time periods [10]. Lastly, indications should be explicitly stated for purposes of creating a consistent patient group [10].

Analyzing Disease Patterns

- Aid in formulating a reinforced hypothesis for future prospective studies
- Provide information regarding natural history, recovery, and prognostic factors
- Postulate relevant measures of interest: sample size, relevant covariates, and length of follow-up

In addition to a reinforced hypothesis, case series can provide viable information for disease patterns regarding the natural history, recovery, and prognostic factors which can further postulate relevant measures of interest including sample size, relevant covariates, and/or length of follow-up [1, 10].

33.3 Reporting Outcomes

By a retrospective approach, ease of accessibility to long-term follow-up and lack of a comparison group make the ideal outcomes primarily

consistent of treatment safety and diagnostic accuracy [10]. Case series are descriptive, and therefore findings should be exclusively presented by descriptive statistics, because comparative tests supported by p-values are irrelevant in the matter and should be avoided, along with conclusive reports; the treatment of interest and its relative efficacy would be supported by an immaterial hypothesis. The variable potential for bias is also important to explicitly state for future prospective studies to ascertain validity of the treatment [10].

The most important outcomes that can be determined by clinical research are measures of physical function and well-being. These are most often deemed significant by their measures of validity and reliability. Validity and reliability refer to the ability of the study to measure the variable of interest and the extent that repeated measures display similar results, respectively. The Western Ontario and McMaster Universities (WOMAC) Osteoarthritis index score is an example of an outcome that correlates well with previously described instruments of pain, stiffness, and physical function [2].

A classification scheme for variable measures of outcomes has previously been proposed by Wilson and Cleary, in which the higher-level outcomes were increasingly influenced by outside, uncontrollable factors of individuals, environments, and/or nonmedical factors that increase the difficulty in measurement and ultimately the definition for the patient population [16]. These five levels of outcomes were suggested as biological and physiological variables (Level I), symptom status (Level II), functional status (Level III), general health perceptions (Level IV), and overall quality of life (Level V) [16]. Instruments for reporting outcomes among orthopedic research can be utilized in many forms including mixed clinician-based and functional outcome; system-specific, disease-specific, and general health-related quality of life; and overall health-related quality of life.

Mixed outcomes occasionally yield clouded findings due to the obstacle of interobserver reliability across physical examinations [14]. Additionally, the broad report of measures into a summarized score deters from the specificity of

patient expectations by range of motion, stiffness, and/or pain, thereby failing to discriminate between improved and worsened patients [13].

System-specific outcomes are localized by a specific body region, while disease-specific outcomes are localized by the specific disease encountered by the patient [13]. Frequently, patients can be evaluated by both approaches. Patients suffering from osteoarthritis of the knee can be assessed by instruments related to well-being of the knee and/or instruments related to well-being of osteoarthritis, increasing the available comparative measures and sensitivity to change [13].

General and overall health-related quality-of-life instruments mainly pertain to the health and happiness, respectively, of a patient to encounter daily activities. The former is a multifactorial concept comprising physical, mental, and social factors that deal with a broad range of activities including work, hobbies, and social interactions; these factors thus measure the patient's ability to carry out everyday life [13]. The latter instrument is similar but primarily concerns the patient's satisfaction with their ability to participate in daily activities [13].

33.4 Level IV Case Series Example

A great example of case series research is a study performed by Geeslin and LaPrade, in which they aimed to report objective stability and subjective outcomes for a prospective series of patients with an acute grade III posterolateral corner (PLC) knee injury treated with anatomic repair and/or reconstruction of all injured structures [6]. At this time, there was a lack of current reports regarding surgical treatment and outcomes of acute PLC injuries, with most of the literature having been published a decade prior [5, 15].

Upon completion of this case series, the authors were able to report significantly improved objective stability, relative to preoperative conditions, that resulted from treatment of grade III PLC injuries by acute repair of avulsed fractures, reconstruction of midsubstance tears, and concurrent reconstruction of any cruciate ligament

tears [6]. This case series adds significant value to the literature in showing the high yield of outcomes by acute repair of the PLC structures.

33.5 Strengths and Pitfalls

Case series, and all observational studies, are uncontrolled, meaning the physician does not choose the treatment for the research participant. This method carries both a positive and negative aspect in the research process. Results of uncontrolled studies more often closer resemble routine clinical practice than randomized trials and can often be better applied to clinical practice [1, 8]. With a diverse range of patients, a high external validity brings greater clinical relevance to differing medical centers and can better represent the population of interest; strict inclusion criteria established by randomized controlled trials severely reduce the extent that findings can be epitomized with common practice [4, 11]. Furthermore, from the financial and ethical standpoint, the lack of comparison and randomization yields a cost-effective study design, and choice treatment by the patient and physician maintains a consistency for common standards of orthopedic practice [10].

However, an absent comparison group does provide limitations. Causal inferences cannot be made, restricting findings to apparent relationships, because the study design lacks an independent variable to differentiate outcomes by treatment protocol rather than to patient characteristics [10]. Analyzing a condition of interest to a control group provides possible correlations for either the presence or degree of exposure to a specific risk factor and collectors may take measurements unaware of the patient characteristics, reducing the risk of bias [9].

Study design for case series studies that lack a comparison group eliminates the potential for confounding bias but increases the potential for selection and measurement bias which can be dependent by the approach of a prospective versus retrospective study. Prospective studies more regularly embody a consistent manner of study design ranging from inclusion and data collection

to patient follow-up measures, and measurement bias may exist with the absence of a standardized protocol [10]. Patients are expected to display worse outcomes when they have become deceased or switched to another hospital, which creates an instance of selection bias [10]. Measurement bias also becomes prevalent, as differing methods of outcome measurements are used in a study [10].

Take-Home Message

- Case series provide a basis for designing further research studies.
- In the instance of unusual occurrences of a disease, case series are effective for designing hypotheses for future prospective studies, although no hypothesis is tested within the case series itself.
- It is important to remember that case series are unable to make comparisons; rather they exclusively report on the outcome findings relative to the treatment of the patients involved in the series through descriptive statistics.

References

1. Audige L, Hanson B, Kopjar B. Issues in the planning and conduct of non-randomised studies. *Injury*. 2006;37(4):340–8. <https://doi.org/10.1016/j.injury.2006.01.026>.
2. Bellamy N. Pain assessment in osteoarthritis: experience with the WOMAC osteoarthritis index. *Semin Arthritis Rheum*. 1989;18(4 Suppl 2):14–7.
3. Carey TS, Boden SD. A critical guide to case series reports. *Spine (Phila Pa 1976)*. 2003;28(15):1631–4. <https://doi.org/10.1097/01.BRS.0000083174.84050.E5>.
4. Dalziel K, Round A, Stein K, Garside R, Castelnuevo E, Payne L. Do the findings of case series studies vary significantly according to methodological characteristics? *Health Technol Assess*. 2005;9(2):iii–v. 1-146.
5. DeLee JC, Riley MB, Rockwood CA Jr. Acute posterolateral rotatory instability of the knee. *Am J Sports Med*. 1983;11(4):199–207. <https://doi.org/10.1177/036354658301100403>.
6. Geeslin AG, LaPrade RF. Outcomes of treatment of acute grade-III isolated and combined posterolateral knee injuries: a prospective case series and surgical technique. *J Bone Joint Surg Am*. 2011;93(18):1672–83. <https://doi.org/10.2106/JBJS.J.01639>.

7. Hanson BP. Designing, conducting and reporting clinical research. A step by step approach. *Injury*. 2006;37(7):583–94. <https://doi.org/10.1016/j.injury.2005.06.051>.
8. Hartz A, Marsh JL. Methodologic issues in observational studies. *Clin Orthop Relat Res*. 2003;(413):33–42. <https://doi.org/10.1097/01.blo.0000079325.41006.95>.
9. Hess DR. Retrospective studies and chart reviews. *Respir Care*. 2004;49(10):1171–4.
10. Kooistra B, Dijkman B, Einhorn TA, Bhandari M. How to design a good case series. *J Bone Joint Surg Am*. 2009;91(Suppl 3):21–6. <https://doi.org/10.2106/JBJS.H.01573>.
11. Lloyd-Williams F, Mair F, Shiels C, Hanratty B, Goldstein P, Beaton S, Capewell S, Lye M, McDonald R, Roberts C, Connelly D. Why are patients in clinical trials of heart failure not like those we see in everyday practice? *J Clin Epidemiol*. 2003;56(12):1157–62.
12. McMaster WC, Sale K, Andersson GB, Bostrom MP, Gebhardt MC, Trippel SB, Clark DC. The conduct of clinical research under the HIPAA Privacy Rule. *J Bone Joint Surg Am*. 2006;88(12):2765–70. <https://doi.org/10.2106/JBJS.F.00794>.
13. Poolman RW, Swiontkowski MF, Fairbank JC, Schemitsch EH, Sprague S, de Vet HC. Outcome instruments: rationale for their use. *J Bone Joint Surg Am*. 2009;91(Suppl 3):41–9. <https://doi.org/10.2106/JBJS.H.01551>.
14. Pynsent P, Fairbank J, Carr A. Outcome measures in orthopaedics and orthopaedic trauma. New York: Oxford University Press; 2004.
15. Stannard JP, Brown SL, Farris RC, McGwin G Jr, Volgas DA. The posterolateral corner of the knee: repair versus reconstruction. *Am J Sports Med*. 2005;33(6):881–8. <https://doi.org/10.1177/0363546504271208>.
16. Wilson IB, Cleary PD. Linking clinical variables with health-related quality of life. A conceptual model of patient outcomes. *JAMA*. 1995;273(1):59–65.



How to Perform a Clinical Study: Level 4 Evidence—Case Report

34

Andrew J. Sheean, Gregory V. Gasbarro,
Nasef M. N. Abedelatif, and Volker Musahl

34.1 The Case for Case Reports

The advent of evidenced-based medicine has revolutionized orthopaedic clinical research. Surgeons, now more than ever, seek answers to diagnostic and therapeutic dilemmas in the pages of medical journals, which have raised the bar for what constitutes a “publishable” study. Study methodology is now routinely scrutinized with various quality assessment tools, and, depending upon the type of study, authors are required to describe their methods using standardized flow diagrams and predetermined checklists. Better methods yield more precise results, which support stronger conclusions. These conclusions are more likely to alter practice habits and improve care. Moreover, authors recognize readers’ (and journal editorial boards’) appetites for higher levels of evidence and now increasingly strive to complete prospective initiatives that routinely compare unique patient cohorts or different treatment strategies. These observations are substantiated by a recent systematic review of available randomized controlled trials (RCT) pertaining to anterior cruciate ligament (ACL) reconstruction, which noted that 60% of the 412 relevant studies

have been published in the last 10 years [2]. Considering the current state of orthopaedic literature, one cannot help but to ask: “Is there still a place for case reports?”

The utility of case reports is derived from the novelty of both musculoskeletal conditions and their treatments. New conditions are discovered, previously described conditions present themselves in unique ways, and modifications to existing therapies or new therapies altogether can enhance patient care. The publication of a well-conceived case report need not be at odds with contemporary evidence-based medicine [1, 6]. Rather, the case report should occupy a special place within the orthopaedic literature. The case report allows its author to alert readers to a novel aspect of medicine, which can improve the clinician’s diagnostic acumen and/or stimulate larger scale initiatives to better understand the benefits of a particular innovation. The purpose of this chapter is to provide a blueprint for the publication of a case report that is focused, informative, and capable of reaching the widest audience possible.

34.2 Case Report Taxonomy

The case report is a retrospective analysis of one, two, or three clinical cases. These types of studies in the orthopaedic literature can generally be classified as one of three types based upon the

A. J. Sheean (✉) · G. V. Gasbarro · V. Musahl
Department of Orthopaedic Surgery,
University of Pittsburgh, Pittsburgh, PA, USA

N. M. N. Abedelatif
Beni Suef University, Beni Suef, Egypt

focus of the report: a novel musculoskeletal condition, a novel presentation of a previously described musculoskeletal condition, or a novel treatment. The following clinical vignettes provide an example of each of the three types of case reports and demonstrate how the authors successfully explain the implications of their observations.

Clinical Vignette 1: A Novel Condition

A 22-year-old collegiate rugby player with a history of recurrent, anterior glenohumeral instability and a large bony Bankart lesion was indicated for bony fixation and capsulorrhaphy [7]. After an unsuccessful attempt at arthroscopic fixation, an arthrotomy and open Bankart repair was performed through a deltopectoral approach. During the exposure, an aberrant muscle belly originating from the long head of the biceps brachii and inserting on the lesser tuberosity, proximal to the subscapularis, was noted. Furthermore, the anterior humeral circumflex vessels were found to lie superficial to the aberrant muscle belly. This variation was recognized, and fixation of the bony Bankart lesion and capsulorrhaphy were completed. The authors emphasized an awareness for this anatomic variation for two reasons. First, exposure and mobilization of the subscapularis was possible without releasing the anterior humeral circumflex vessels. Second, adequate visualization and management of the subscapularis was contingent upon the intraoperative recognition of the aberrant muscle belly.

Clinical Vignette 2: A Novel Presentation of a Previously Described Musculoskeletal Condition

A 32-year-old man presented to the emergency department with involuntary loss of bowel and bladder function with tetraparesis

after sustaining a hyperextension-flexion injury of the neck while riding a roller coaster [3]. Magnetic resonance imaging (MRI) of the cervical spine demonstrated a synovial cyst associated with an os odontoideum causing severe spinal stenosis. The patient subsequently underwent cyst needle aspiration of the synovial cyst, C2–C3 decompression, laminectomy, and instrumented posterior spinal fusion from C1 to C3. Synovial cysts in the spine commonly are found in the lumbar spine, and the authors pointed out the rarity of the synovial cyst location at the atlanto-axial junction in this case. The authors emphasized an awareness for the possibility of such a lesion found in association with an os odontoideum and recommended that those patients with os odontoideum be screened for atlanto-axial instability and for subtle clinical findings suggestive of a cervical myelopathy.

Clinical Vignette 3: A Novel Treatment

A 26-year-old woman with a history of two anterior hip dislocations following an arthroscopic labral repair was treated with revision hip arthroscopy and capsular plication before experiencing a third anterior hip dislocation while jogging [4]. After a thorough radiographic work-up, her hip instability was attributed to the combined effect of relative acetabular anteversion and focal anterior acetabular deficiency. She subsequently underwent a periacetabular osteotomy (PAO) to negate the effects of what was described as an overzealous anterior acetabuloplasty. At 1 year postoperative, the patient had not experienced any further anterior hip dislocations and demonstrated markedly improved patient-reported outcomes (PRO). The authors emphasized the importance of a thorough

preoperative assessment of acetabular version prior to undertaking arthroscopic acetabuloplasty and described the utility of PAO to effectively treat iatrogenic anterior hip instability.

34.3 The Process

The process of writing a successful case report should proceed in a stepwise fashion (Table 34.1) with a review of the pertinent literature after the recognition of a unique clinical scenario. Once it has been confirmed that the clinical scenario is of significant novelty to warrant further efforts, an appraisal of the medical record should be undertaken to confirm that all relevant details are available and of sufficient quality to be included in a

Table 34.1 Steps for writing a successful case report

manuscript suitable for publication. Incomplete or missing clinical notes, undocumented laboratory data, and/or imaging studies of poor quality can all significantly hamper the process of presenting a case in an informative and thorough manner. For case reports describing a novel treatment approach, clinical follow-up of at least 1 year is preferable, and the inclusion of pre- and post-treatment patient-reported outcomes (PRO) strengthens conclusions pertaining to utility of the proposed treatment. The patient's consent should be obtained in accordance with local institutional review board standards. Furthermore, proof of consent is frequently required by journals considering case reports for publication.

Prior to writing the case report, it is advisable to select a target journal to which the manuscript will be submitted. This should be done based upon the relevance of the case report to the subject matter typically published by a particular journal. The author should ask himself or herself: “Would this case report be of interest to this journal’s readership?” Additionally, case reports are not uniformly accepted for publication by all journals, and it is important to select a target journal based upon a precedent for the publication of case reports. Fact Box 34.1 provides a synopsis of the historical record of a collection of orthopaedic journals for publishing case reports.

Fact Box 34.1: Last Published Case Report in Orthopaedic Subspecialty Journals. Impact Factor According to the 2016 Journal Citation Report (Clarivate Analytics)

Orthopaedic journal name	Impact factor	Last published case report
<i>American Journal of Sports Medicine</i>	5.673	April 2014
<i>Journal of Bone and Joint Surgery</i>	4.840	June 2012
<i>Arthroscopy: Journal of Arthroscopic and Related Surgery</i>	4.292	January 2012
<i>Knee Surgery, Sports Traumatology and Arthroscopy</i>	3.227	June 2017
<i>Journal of Arthroplasty</i>	3.055	March 2013
<i>The Spine Journal</i>	2.962	November 2017
<i>Bone and Joint Journal</i>	2.953	July 2010
<i>The Journal of the American Academy of Orthopaedic Surgeons</i>	2.782	August 2017
<i>Clinical Orthopaedics and Related Research</i>	2.765	February 2016
<i>Journal of Shoulder and Elbow Surgery</i>	2.730	December 2017
<i>Journal of Orthopaedic Trauma</i>	2.251	June 2015
<i>Foot and Ankle International</i>	1.872	February 2017
<i>Journal of Pediatric Orthopaedics</i>	1.695	September 2017
<i>Journal of Hand Surgery</i>	1.606	August 2017

34.4 The Structure

Brevity is a critical feature of a successful case reports, and the final manuscript should be a distillation of the key features of a novel clinical scenario. Limitations on word count, number of figures, and number of references vary somewhat from journal to journal, which underscores the importance of a complete review of manuscript guidelines prior to beginning to write (Fact Box 34.2). Case reports are not uncommonly limited to between 1000 and 1500 words [1]. While the composition of the case report varies based upon the target journal, the general structure follows a relatively consistent format (Table 34.2.).

Table 34.2 Key components of a case report

Abstract	<ul style="list-style-type: none"> – Summarize the clinical scenario – State the one or two purposes of the case report
Introduction	<ul style="list-style-type: none"> – Provide a context for the case report with a brief synopsis of the relevant literature – Describe the novel aspects of the clinical scenario – State the one or two purposes of the case report
Case presentation	<ul style="list-style-type: none"> – Provide chronologic description of case details – Catalog pertinent physical examination details, laboratory values, and imaging results – Describe clinical and radiographic follow-up
Discussion	<ul style="list-style-type: none"> – Summarize key observations – Describe the novel aspects of the clinical scenario – Review the relevant literature – Provide concluding remarks that restate the one or two purposes of the case report

Fact Box 34.2: Example of Author Instructions for Journals That Have Published a Case Report After January 2017 and Still Currently Accept Submissions

Orthopaedic journal name	Instructions for author
<i>Knee Surgery, Sports Traumatology and Arthroscopy</i>	<ul style="list-style-type: none"> • Short description of one (or few) case(s) that are more or less unique in their presentation and/or a new unique form of treatment • Please note that we only publish unique and rare case reports and generally discourage them • Specifications: 100-word abstract, 1000-word text including references • http://www.kssta.org/authors/instructions/casereport/
<i>The Journal of the American Academy of Orthopaedic Surgeons</i>	<ul style="list-style-type: none"> • Pre-submission approval with a proposal or formal invitation is not required • Summarize pertinent unusual or unexpected elements of case with discussion reviewing scientific/educational value • Informed consent must be obtained from participating subject(s) • Specifications: 150-word abstract, 1500-word text, 3 figure panels, and 10 references • http://edmgr.ovid.com/jaaos/accounts/authinst.pdf
<i>Journal of Shoulder and Elbow Surgery</i>	<ul style="list-style-type: none"> • Encourage submission to JSES open access, a quarterly, online publication • \$750 submission fee • Minimum 2-year follow-up is not required • Specifications: No abstract, include keywords at the end of the introduction, and 2250-word text • http://www.jshoulderelbow.org/content/authorinfo
<i>Foot and Ankle International</i>	<ul style="list-style-type: none"> • <i>Very few case reports are accepted for publication</i> • Case reports must offer either new information that has been previously unpublished and offer completely new information or information that will change the current practice patterns of our readers • Entities that are unique in and of themselves bizarre, or common, will <i>not</i> be accepted as case reports • Sections should include introduction, case report, discussion, and summary/conclusion • Specifications: no abstract • https://us.sagepub.com/en-us/nam/journal/foot-ankle-international#CaseReports

Fact Box 34.2: (continued)

Orthopaedic journal name	Instructions for author
<i>Journal of Hand Surgery</i>	<ul style="list-style-type: none"> • To be worthy of publication, a case report must have extraordinary teaching value to the readers • Typically, cases where two findings are associated are not accepted since the findings are often coincidentally rather than causally related • Sections should include introduction, case report, and discussion • Specifications: one-paragraph description of manuscript contents, 150-word abstract, 1500-word text, and 10 references • http://www.jhandsurg.org/content/authorinfo

34.4.1 Title

The title should be concise and informative. Keywords relevant to the clinical scenario should be included so as to maximize the likelihood of being captured in subsequent literature searches.

34.4.2 Abstract

If allowed, abstracts are generally limited to between 150 and 200 words and should provide a snapshot of clinical scenario. The emphasis of the abstract should be on defining the novelty of the case so as to capture the readers' interest. The abstract should include the one or two "take-home messages" of the case report.

34.4.3 Introduction

Although not uniformly required, the introduction provides the author with an opportunity to elaborate on what the abstract has already stated in terms of the novelty of the clinical scenario. The introduction should clearly explain *what* is unique about the clinical scenario being presented and *why* the case report is worthy of the readers' interest. The focus of the introduction should be on efficiency as most of the manuscript should be comprised of the case presentation and the discussion sections. References to prior relevant studies should only be made to underscore the unique features of the clinical scenario

described, and a more thorough review of the existing literature can be reserved for the discussion section.

34.4.4 Case Presentation

The case should be presented in chronological order, beginning with the patient's initial presentation. The patient's subjective complaints, components of the physical examination, pertinent laboratory values, and imaging studies should be provided devoid of the authors' analysis and/or inferences so as to allow the reader to establish his or her own conclusions about the case's validity [5]. Given the frequent restrictions on figures and/or graphics, only the most relevant clinical photographs, radiographs, pathology specimens, or advanced imaging results should be included for review and be of the requisite quality to allow for the readers' thorough interpretation. For case reports detailing a novel treatment or surgical technique, a thorough account of the operative technique should be included. Here again, clinical photographs and diagrams may be of particular utility to clearly explain unique intraoperative findings or a novel surgical technique.

The importance of adequate post-treatment follow-up cannot be overstated for case reports presenting novel treatment approaches. The success of a treatment is best substantiated by PRO and/or imaging results obtained at least 1 year post-treatment. Without adequate data, it is difficult, if not impossible, to assert the utility of a particular treatment. Moreover, case reports

espousing the utility of novel treatments without sufficient clinical follow-up are unlikely to compel readers to seriously consider the implementation of the proposed therapy.

34.5 Discussion

The discussion section summarizes the key aspects of the clinical scenario that explain its worthiness of a case report. For anatomic variations, “normal” should be defined based upon prior epidemiologic descriptions, and the implications of the variation should be discussed. For entirely new conditions, the clinical presentation should be detailed and possible treatment strategies proposed. Novel presentations of known musculoskeletal conditions should be discussed in relation to the modalities typically used to make the diagnosis. Was there something unique about the current case that obscured the diagnosis, and, if so, how was the diagnosis made? For new treatments, the rationale for a unique approach should be explained. Does the new treatment address deficiencies in more traditional therapies? Additionally, the proposed advantages of the new treatment should be clearly described.

In each of these circumstances, a focused review of the relevant literature is critical in order to place the case report in a broader context. The literature review serves the purpose of explaining how the case report differs from what has already been described. However, only the most pertinent references should be cited, as most case reports are limited to including between 10 and 15 references.

34.6 Conclusion

The publication of case reports facilitates the discussion of novel musculoskeletal conditions,

clinical presentations, and treatment approaches. Through a concise, focused description of a unique clinical scenario, authors can employ the case report to share important details about diagnosis and treatment. The dissemination of this information has the potential to sharpen diagnostic practices and catalyze treatment innovation. In these ways, the case report remains a worthwhile addition to the orthopaedic literature.

Take-Home Message

- In the era of evidence-based medicine, the importance of case reports stems from their utility in describing new conditions, communicating a novel presentation of a previously described conditions, or detailing new treatment approaches.

References

1. Friedman JN. The case for ... writing case reports. *Paediatr Child Health*. 2006;11:343–4.
2. Kay J, Memon M, Sa D, Simunovic N, Musahl V, Fu FH, et al. A historical analysis of randomized controlled trials in anterior cruciate ligament surgery. *J Bone Joint Surg Am*. 2017;99:2062–8.
3. Schmitz MR, Jenne J. Acute tetraparesis caused by a cervical spine synovial cyst associated with an os odontoideum: a case report. *JBJS Case Connect*. 2012;2:e17.
4. Sheean AJ, Barrow AE, Burns TC, Schmitz MR. Iatrogenic hip instability treated with periacetabular osteotomy. *J Am Acad Orthop Surg*. 2017;25:594–9.
5. Sun Z. Tips for writing a case report for the novice author. *J Med Radiat Sci*. 2013;60:108–13.
6. Vandenbroucke JP. In defense of case reports and case series. *Ann Intern Med*. 2001;134:330–4.
7. Warner JJ, Paletta GA, Warren RF. Accessory head of the biceps brachii. Case report demonstrating clinical relevance. *Clin Orthop Relat Res*. 1992;179:181.



Level 5: Evidence

35

Seán Mc Auliffe and Pieter D'Hooghe

Fact Box 35.1

Level 5 evidence within the levels of evidence framework refers to the use of opinions of respected authorities, consensus statements, descriptive studies, or reports of expert committees involving information “without explicit critical appraisal or based on economic theory or first principles” (Oxford Centre for Evidence-based Medicine Levels of Evidence, May 2001).

Fact Box 35.2

Expert consensus is seen as a suitable method to validate hierarchies of evidence or indeed other components of the evidence-based medicine enterprise, such as the CONSORT Statement for reporting randomized controlled trials and principles for developing practice guidelines.

Fact Box 35.3

The Delphi technique involves the gathering of expert opinion by means of structured or semi-structured rounds of questioning subsequently leading to a consensus of opinion on a particular topic.

35.1 Introduction

Clinicians are often faced with difficult decisions and uncertainty when patients need a certain treatment. They routinely rely on scientific literature in addition to their knowledge, experience, and patient preferences to make informed decisions. However, what if no such scientific literature exists, and what if there are no available RCTs or cohort studies in order to aid their clinical decisions? Therefore, level 5 evidence methods provide a necessary and important starting point for clinicians where other, more robust evidence is unavailable, overcoming some of the limitations within evidence-based medicine. In essence level 5 evidence methods embody the principle of practice-based evidence by providing results that are relevant to the local population and culture and are readily implementable within the healthcare system. This approach fits with the call for more “practice-based evidence,”

S. Mc Auliffe · P. D'Hooghe (✉)
Aspetar Orthopaedic and Sports Medicine Hospital,
Doha, Qatar

Aspire Zone, Doha, Qatar
e-mail: sean.auliffe@aspetar.com;
pieter.dhooghe@aspetar.com

where the evidence is gathered in real-life clinical settings and there is greater emphasis on the external validity of the evidence (generalizability) rather than on its internal validity (validity of causal inference) in order to develop better interventions [1, 2].

Despite the obvious benefit of expert consensus and guidelines, these methods are not without their limitations.

Firstly, the methods of developing expert opinions or consensus statements are often not reported clearly, and thus, one cannot be certain how the evidence has been collected or assessed. For example, the recent “World Heart Expert Consensus Statement on Antiplatelet Therapy in East Asian Patients with acute coronary syndrome (ACS) or undergoing percutaneous coronary intervention (PCI)” summarized the latest data on the use of antiplatelet in the East Asian populations with ACS or for whom PCI is performed [3]. An expert panel of a wide array of experiences and knowledge writes the document, and yet, the methodology of the process summarizing the evidence cited in this statement remains unreported [4].

A second limitation of level 5 evidence is that recommendations and consensus statements are influenced by the opinions, clinical experience, and composition of the group that develops them. The beliefs and opinions to which experts subscribe can be based on misconceptions and personal bias recollections that misrepresent population norms or may not reflect the view of their respective profession as a whole [5, 6].

Finally in contrast to a formal evidence-based medicine (EBM) approach, consensus statements and guidelines may be unduly influenced by ascertainment bias (inclusion of subjects not representative of the general population or selective analysis of groups at opposite extremes) or an emotional case study (more recent or more publicized stories or successful patient encounters) to influence the final recommendations. Hence higher-level evidence-based methodologies are usually free of such influences attributed to requirements and guidelines for eligibility

criteria, protocol registration, and classification of research quality to name a few [7].

Considering the obvious limitations of level 5 evidence-based methods, we can recommend the utilization of these methods in contributing to informed clinical decision-making and informing clinical practice. An important counter-question to ask is: How were some types of evidence assigned to be higher in the hierarchy than others? The answer is “expert consensus.” Expert consensus is seen as a suitable method to validate hierarchies of evidence or indeed other components of the evidence-based medicine enterprise, such as the CONSORT Statement for reporting randomized controlled trials and principles for developing practice guidelines [8, 9]. More generally, consensus has an important role in the scientific process. New theories gain ground as more members of the scientific community see a new theory as giving a better account of the evidence than older ones and thus act as a fundamental starting point in the evidence-based medicine approach.

35.2 Consensus Group Methods: Definition and Rationale

Level 5 evidence-based methods may be achieved through many formal methods, the most common of which are expert opinions or consensus group methods such as the Delphi or nominal group technique (NGT) [2]. The goal of these methodologies is to establish how well experts agree on a particular topic or issue, with the premise that accurate and reliable assessment can be best achieved by consulting a panel of experts and generating a group consensus on that particular issue. Consensus group methods have demonstrated acceptable construct validity and reliability [6, 10]. Utilizing these methods may assist in the development of an evidence base that can guide decisions, policy makers, and practitioners, thus ensuring practitioners can move beyond relying on their own individual experiences by drawing on the accumulated

body of knowledge of a larger, expert group [2]. Common features necessary for carrying out consensus group methods include anonymity, iteration, controlled feedback, statistical group response, and structured interaction all of which differentiate informal group meetings from formal consensus methods [11]. Consensus techniques such as the NGT and Delphi technique are similar to focus groups. A key strength of consensus methods is the balanced participation from group members, unlike a focus group, whereby the facilitator must control for, and minimize the risk of, a dominant participant influencing the discussion [12].

35.3 The Delphi Method

The Delphi technique involves the gathering of expert opinion by means of structured or semi-structured rounds of questioning subsequently leading to a consensus of opinion on a particular topic [13]. The Rand Corporation first developed the Delphi technique in 1953 [14]. There is no standard method to calculate a panel size for the Delphi technique with a sample of about 15 commonly accepted in the literature [12].

Carrying out a Delphi study involves a series of steps and choices. The first step in the Delphi technique involves a sufficient review of the available literature in order to generate and clarify a clear research question. The next step involves the selection of a suitable expert panel or individuals with expertise relevant to the question. Following the formation of a suitable panel of individuals, a questionnaire is constructed by the coordinator, which will then be used to gather opinions from the expert panel.

The expert panel is then provided with an initial questionnaire (round one) that uses open-ended questions to generate a list of ideas or concepts related to the research question. The first-round questionnaire will present a series of statements that the respondent is asked to rate on a clearly defined Likert scale. Often a nine-point Likert scale is used for the rating, although three-

point, five-point, and seven-point scales have also been used [14–17]. Respondents are asked, to rate both items, as well as to compose free-text comments that, for example, explain their rating and express disagreement with the statement's relevance [18]. While systematic literature searches and focus groups are often used for the initial Delphi questionnaire, expert panel members are given the opportunity to suggest additional items when completing the questionnaire [19].

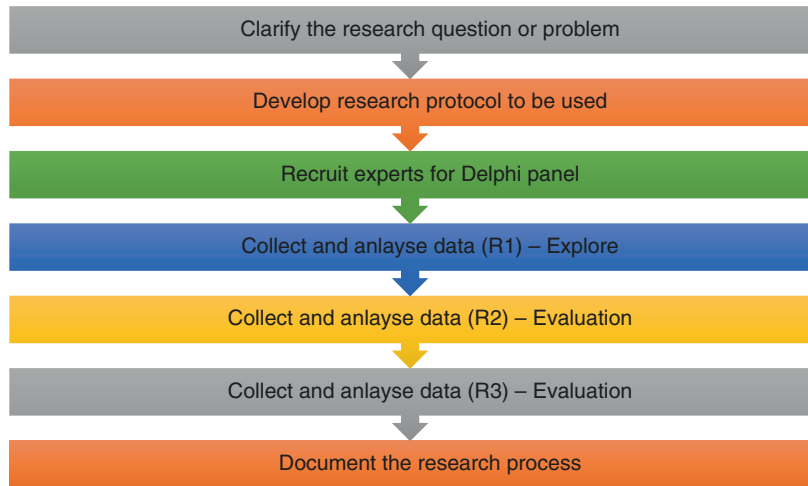
Members of the research team analyze, collate, and compile an inclusive list of responses for subsequent resubmission to the expert panel in the form of a second-round questionnaire. In all subsequent rounds of the Delphi process, the panel members are provided with feedback pertaining to the individual responses provided and those of the other panelists. The process of questionnaire circulation and controlled opinion feedback is continued until consensus has been reached. Ordinarily the Delphi process requires between two and four rounds to gather a consensus of judgment, opinion, or choice.

The results of the multiple rounds of a Delphi study may be difficult to communicate due to the complexity. One common method involves the use of flowcharts showing the fate of items at each round. The simplest output method from a Delphi study is a list of accepted items or alternatively where the number of accepted items is small, simply reporting the list may in fact be sufficient [19].

Some methodological concerns regarding the Delphi method that should be taken under consideration before carrying out such a study include the definition of an expert panel member, the potential bias of selecting panel members, anonymity of individual responses, questionnaire design, and scoring methods and, therefore, should all be considered and addressed prior to undertaking the study [20].

A graphical representation of the multiple stages involved in the Delphi process is outlined in Fig. 35.1.

Fig. 35.1 Delphi study protocol



35.4 Nominal Group Technique (NGT)

The NGT method incorporates a structured face-to-face group interaction. The NGT involves a four-stage process, namely, silent generation, round robin, clarification, and voting (ranking or rating) [21].

To initiate the NGT process, one to two questions are sent to participants in advance of a face-to-face group meeting. Subsequently, at the beginning of the meeting, participants are given time to reflect or record their individual ideas in response to a question, in a process known as “silent generation.” The next stage of the NGT process is known as the “round robin” process, in which the facilitator requests one participant at a time to state a single idea to the group. Participants are afforded as much time as required until no new further ideas are generated, with ideas recorded verbatim on, for example, a flipchart or white board. The third stage of the NGT process is known as the “clarification stage” [21]. This stage provides the opportunity for clarification of the ideas generated in previ-

ous rounds, as well as providing ample opportunity for a grouping step, where similar ideas are grouped together with agreement from all participants. Participants are also afforded the opportunity to exclude, include, or alter ideas, as well as generate grouping themes. Both the round robin and clarification phases may take up to 30 min, respectively [12].

The next stage of the NGT method is known as the “voting or ranking” stage where participants are provided with a ranking sheet, requiring them to identify their top preferences from the generated ideas in the previous stages with larger numbers reflecting greater importance. The number of items chosen by participants may vary according to the topic of interest, but the ranking of five ideas is common in the literature [12, 22–27]. The final stage of the NGT process is referred to as the “discussion stage,” where the scores for each idea are summed and presented to the group for discussion. The timing for this stage is likely to depend on a number of factors, including the complexity of the topic and how many items need to be prioritized. A graphical illustration of the NGT method is outlined in Fig. 35.2 [12].

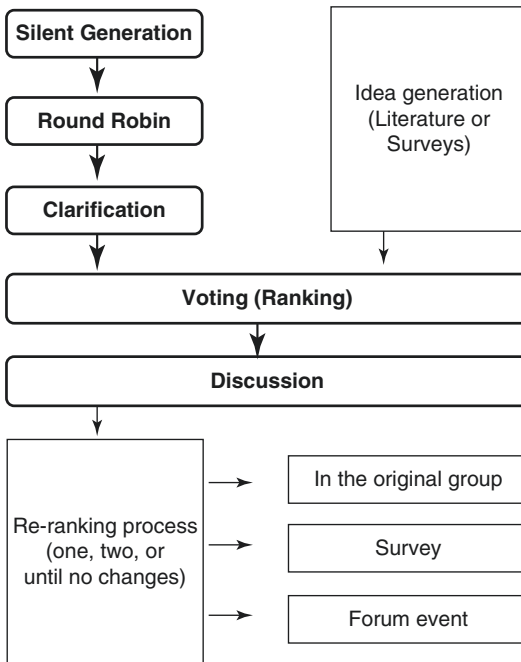


Fig. 35.2 NGT method

Clinical Vignette: The Practical Implementation of Level 5 Evidence in Clinical Orthopedic Medicine

To illustrate the use of expert consensus research methods in orthopedic medicine, the International Society on Cartilage Repair of the Ankle recently held a consensus meeting in Pittsburgh, Pennsylvania, USA, in order to develop a consensus opinion on key topics within cartilage repair of the ankle. Using the Delphi consensus group method described previously, this consensus meeting assembled orthopedic surgeons, physical therapists, radiologists, and basic scientists to provide evidence-based and/or expert recommendations in the field on cartilage repair of the ankle.

The consensus meeting which took place on the 17th–18th of November 2017 at the University of Pittsburgh (Fig. 35.3) represents a culmination of the Delphi process methodology which began in early 2017 and ultimately aims to help clinicians provide appropriate treatment of ankle articular cartilage injuries based on expert-level opinion.

35.5 Conclusion


In essence, level 5 evidence methods embody the principle of practice-based evidence by providing results that are relevant to the local population and culture and are readily implementable within the healthcare system. This approach fits with the call for more “practice-based evidence,” where the evidence is gathered in real-life clinical settings and there is greater emphasis on the external validity of the evidence (generalizability) rather than on its internal validity (validity of causal inference) in order to develop better interventions.

Take-Home Message



- By using expert opinion and consensus statements to develop an evidence base, practitioners can move beyond relying on their own experience and can avail on an accumulated experience of a larger cohort of practitioners in their respective fields.
- Level 5 evidence may also facilitate the development of clinical practice guidelines or in assisting healthcare stakeholders with decision making, thus highlighting opportunities and areas for future research.
- Although at the lower end of the evidence-based medicine hierarchy, level 5 evidence methods have the added advantage of being feasible with limited resources where it may not be feasible to carry out randomized controlled trials, population surveys, or cohort studies.

INTERNATIONAL CONSENSUS MEETING ON CARTILAGE REPAIR OF THE ANKLE

SAVE THE DATE



November 17 and 18, 2017
Pittsburgh, PA, USA

This meeting represents the culmination of an intense, year-long collaboration among global leaders in cartilage repair of the ankle to develop consensus on key focus areas within the field. This first-of-its-kind effort has assembled orthopaedic surgeons, physical therapists, radiologists, and basic scientists to provide evidence-based and/or expert recommendations on the diagnosis, treatment, rehabilitation, and follow-up for a common and difficult clinical problem.

You're invited to attend the International Consensus Meeting on Cartilage Repair of the Ankle. You will have the opportunity to witness the consensus discussion and rigorous scientific debate, as well as to network with national and international leaders in the field in an informal setting.


ORGANIZING TEAM

James D. F. Calder, MD – London, UK	Jin Woo Lee, MD, PhD – Seoul, Korea
Pieter D'Hooghe, MD – Doha, Qatar	Helder Pereira, MD – Vila do Conde, Portugal
Richard D. Ferkel, MD – Los Angeles, USA	James W. Stone, MD – Milwaukee, USA
John G. Kennedy, MD – New York, USA	Masato Takao, MD, PhD – Tokyo, Japan

PITTSBURGH HOSTS

MaCalus V. Hogan, MD	Christopher D. Murawski
----------------------	-------------------------

WHERE:
University Club • 123 University Place • Pittsburgh, PA 15260



International Consensus Meeting on Cartilage Repair of the Ankle
November 17-18, 2017
Pittsburgh, Pennsylvania, USA

Fig. 35.3 A practical example on the Delphi method used at a surgical meeting in Pittsburgh on cartilage repair of the ankle

35.6 Website

- Oxford Centre for Evidence-based Medicine Levels of Evidence (May 2001). <http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/>

References

1. Green LW. Making research relevant: if it is an evidence-based practice, where's the practice-based evidence? *Fam Pract.* 2008;25(Suppl 1):20–4.
2. Minas H, Jorm AFO. Where there is no evidence: use of expert consensus methods to fill the evidence gap in low-income countries and cultural minorities. *Int J Ment Health Syst.* 2010;4:33. <https://doi.org/10.1186/1752-4458-4-33>.
3. Levine G, Jeong Y, Goto S, Anderson J, Huo Y, Mega J, Taubert K, Smith S. Expert consensus document: World Heart Federation expert consensus statement on antiplatelet therapy in East Asian patients with ACS or undergoing PCI. *Nat Rev Cardiol.* 2014;11(10):597–606. <https://doi.org/10.1038/nrcardio.2014.104>. Epub 2014 Aug 26.
4. Kwong JSW, Chen H, Sun X. Development of evidence-based recommendations: implications for preparing expert consensus statements. *Chin Med J (Engl).* 2016;129:2998–3000.
5. Kane RL. Creating practice guidelines: the dangers of over-reliance on expert judgment. *J Law Med Ethics.* 1995;23:62–4.
6. Pacini D, Murana G, Leone A, Di Marco L, Pantaleo A. The value and limitations of guidelines, expert consensus, and registries on the management of patients with thoracic aortic disease. *Korean J Thorac Cardiovasc Surg.* 2016;49(6):413–20. <https://doi.org/10.5090/kjtc.2016.49.6.413>.
7. Echemendia RJ, Giza CC, Kutcher JS. Developing guidelines for return to play: consensus and evidence-based approaches. *Brain Inj.* 2015;29(2):185–94.
8. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomized trials. *BMJ.* 2010;340:32.
9. Shekelle PG, Woolf SH, Eccles M, Grimshaw J. Developing clinical guidelines. *West J Med.* 1999;170(6):348–51.
10. Hutchings A, Raine R, Sanderson C, Black N. A comparison of formal consensus methods used for developing clinical guidelines. *J Health Serv Res Policy.* 2006;11(4):218–24.
11. Jones J, Hunter D. Consensus methods for medical and health services research. *BMJ.* 1995;311:376–80.
12. McMillan SS, Kelly F, Sav A, Kendall E, King MA, Whitty JA, et al. Using the nominal group technique: how to analyse across multiple groups. *Health Serv Outcomes Res Method.* 2014;14:92–108.
13. O'Neill S, Watson PJ, Barry S. A Delphi study of risk factors for achilles tendinopathy- opinions of world tendon experts. *Int J Sports Phys Ther.* 2016;11(5):684–97.
14. Linstone HA, Turoff M. *The Delphi survey: method techniques and applications.* Reading: Addison-Wesley; 1975.
15. Cantrill JA, Sibbald B, Buetow S. Indicators of the appropriateness of long term prescribing in general practice in the United Kingdom: consensus development, face and content validity, feasibility and reliability. *Qual Health Care.* 1998;7:130–5.
16. Cassar Flores A, Marshall S, Cordina M. Use of the Delphi technique to determine safety features to be included in a neonatal and paediatric prescription chart. *Int J Clin Pharmacol.* 2014;36(6):1179–89.
17. Chan A, Tan SH, Wong CM, Yap KY, Ko Y. Clinically significant drug–drug interactions between oral anticancer agents and nonanticancer agents: a Delphi survey of oncology pharmacists. *Clin Ther.* 2009;31(Pt 2):2379–86; 59(4):334–41.
18. McMillan SS, King M, Tully MP. How to use the nominal group and Delphi techniques. *Int J Clin Pharmacol.* 2016;38:655.
19. Jorm A. Using the Delphi expert consensus method in mental health research. *Aust N Z J Psychiatry.* 2015;49(10):887–97.
20. Thangaratinam S, Redman CW. The Delphi technique. *Obstet Gynaecol.* 2005;7:120–5. <https://doi.org/10.1576/toag.7.2.120.27071>.
21. Delbecq AL, van de Ven AH, Gustafson DH. *group techniques for program planning, a guide to nominal group and Delphi processes.* Scott, Foresman and Company: Glenview; 1975.
22. Aljamal M, Ashcroft DM, Tully MP. Development of indicators to assess the quality of medicines reconciliation at hospital admission: an e-Delphi study. *Int J Pharm Pract.* 2016;24(3):209–16.
23. Cross H. Consensus methods: a bridge between clinical reasoning and clinical research? *Int J Lepr Other Mycobact Dis.* 2005;73(1):28–32.
24. Dening KH, Jones L, Sampson EL. Preferences for end-of-life care: a nominal group study of people with dementia and their family carers. *Palliat Med.* 2012;27(5):409–17.
25. Kuhn TS. *The structure of scientific revolutions.* 2nd ed. Chicago: University of Chicago Press; 1970.
26. Sandrey MA, Bulger SM. The Delphi Method: an approach for facilitating evidence based practice in athletic training. *Athl Train Educ J.* 2008;3(4):135–42.
27. Humphrey-Murto S, Varpio L, Gonsalves C, Wood TJ. Using consensus group methods such as Delphi and Nominal Group in medical education research. *Medical Teacher;* 2016.

Part VI

How to Perform a Review Article?



Type of Review and How to Get Started

36

Matthew Skelly, Andrew Duong,
Nicole Simunovic, and Olufemi R. Ayeni

36.1 Introduction

The advent of evidence-based medicine has greatly improved the quantity and quality of research published today. However, with this increased output of high quality research, academics and clinicians are inundated with new reports that require time and resources to appropriately read and review [1, 10, 11]. This might result in many relevant research articles going unread.

One answer to keeping abreast of this increase in literature is the synthesis of information in the form of a review. Reviews attempt to summarize and present the available literature on a given topic. In doing so, they establish whether treatment effects are consistent across studies, strengthen the power and precision of estimated treatment effects and eliminate biases that can be associated with individual studies. This provides healthcare professionals and policy makers a basis upon which they may make evidence-based decisions. Reviews also take stock of what infor-

mation is available and can be useful in identifying gaps in our current knowledge as well as potential avenues for future research.

36.2 Types of Review Articles

There are two main types of review articles: systematic reviews and narrative reviews. Both types aim to abstract information from available resources to answer research questions but differ in their approach.

Systematic reviews follow a planned and reproducible process of searching and identifying relevant articles to answer a proposed question. For example, in adults undergoing a primary hip arthroscopy is the supine or lateral approach more effective at lowering post-operative pain, narcotic usage and improving hip function at 90 days following surgery [13]. They are explicit as to what types of resources they include; the exact procedures by which the primary studies were searched, screened and data abstracted; and how their findings were reached [17]. It is often recommended that researchers design their research questions with the PICO framework (Population, Intervention, Control, Outcome). For the example above, the population is adults undergoing primary hip arthroscopy, the intervention and control are the supine and lateral approach, and the outcomes include post-operative pain, narcotic usage and hip function at 90 days.

M. Skelly · A. Duong · N. Simunovic · O. R. Ayeni (✉)
Department of Surgery, McMaster University,
1200 Main St W, 4E15, Hamilton, ON, Canada
e-mail: duonga@mcmaster.ca; simunon@mcmaster.ca

Narrative reviews, also known as literature reviews, provide an overview of the current state of knowledge on a given topic. For example, a narrative review may look at the broad topic of humeral fractures. They do not necessarily follow a systematic methodology and typically focus on key articles or even expert opinions in a given topic area. Narrative reviews sometimes use non-peer reviewed sources such as editorials, book chapters and interviews. Narrative reviews are not required to describe the methodology by which authors assembled the literature cited in the review, where findings can be heavily dependent on the literature that authors chose to include. This might introduce a significant selection bias and creates the potential for independent authors to arrive at different conclusions, despite seeking to answer the same research question.

- Systematic reviews follow a stringent, planned and reproducible process of searching and identifying relevant articles to answer their proposed research question.
- Narrative reviews do not necessarily follow a systematic methodology and typically focus on key articles or even expert opinions in a given topic area.

36.2.1 Practical Example

In 1998, a narrative review was performed to determine what patient-related factors affect the functional outcome of total hip arthroplasty [19]. This review conducted a brief literature search of one database and included other additional articles identified as relevant by the authors. The authors concluded that the best functional outcomes were reported by patients between the ages of 45 and 75, who weighed less than 70 kg and who had a better preoperative functional status, with few to no baseline comorbidities [19]. The review also indicated that women had better functional outcomes and prosthesis survival rates than men but stated that this may be the result of

confounding factors [19]. In 2004 a systematic review was published on the same topic. It specifically explored factors impacting the health-related quality of life of patients undergoing total hip and knee arthroplasty [5]. It not only examined which patient group had the best functional outcomes but also took into account confounding factors. Among its conclusions were that patient age was not an obstacle to effective surgery, men improved more than women from these total arthroplasties, and after a follow-up time of 1 year, there was no difference in health-related quality of life between weight groups [5]. This helped to dispel prior thoughts of refusing to perform hip arthroplasty on obese patients on the basis of weight [5]. The systematic review also used its data to examine the effect of comorbidities, levels of preoperative function, wait time for surgery and procedure type [5]. The difference in findings and the extent to which each factor was explored can be directly attributed to methodological and source differences between narrative and systematic reviews [8].

Overall it could be said that systematic reviews fall on the more objective end of the spectrum, while narrative reviews are often more subjective [8, 16, 17].

36.3 Why Conduct a Narrative Review over a Systematic Review?

Narrative reviews can be written relatively quickly and provide readers with current knowledge about a certain topic. They tend to be written by authors who are experts in the field and are therefore able to elaborate on their conclusions through their own personal experiences, theories or models and educated opinions. This additional insight can be invaluable in new areas of research that are lacking a sufficient body of literature. In comparison, systematic reviews are time-intensive to perform, making them most useful when there is a large body of primary research studies available to address a specific research question (Table 36.1). In the event that the reviewed studies share a common outcome, a

Table 36.1 A comparison of narrative reviews and systematic reviews

Feature	Narrative review	Systematic review
Research question	Broad overview of a topic	Specific well-defined research question examining a specific aspect of a greater topic
Searching for studies	Searches are not exhaustive and do not guarantee capture of all available literature	Attempt to capture all available published literature and work in progress in a well-documented process. Based on a predefined protocol
Study selection	Reasons for the inclusion/exclusion of studies are not required and are not commonly explained	Reasons for the inclusion/exclusion of studies are explicit and geared towards the research question
Assessment of the quality of included studies	Quality of included studies is not typically evaluated	Systematic assessment of quality of all included studies using established scales, tools, or guidelines
Interpretation and conclusions	Based in part on the resources gathered and in part on the author's intuition/opinion	Based solely on the data gathered

meta-analysis may be performed. A meta-analysis involves the careful consideration of the quality and strength of each study included in the systematic review to create a larger and more accurate pooled estimate of the effect of a treatment.

- Narrative reviews can be written quickly, they tend to be written by experts of the field they concern, and they are invaluable in new areas of research with little available research.
- Systematic reviews are more appropriate when there is a large body of evidence that could benefit from summarization or pooling of results to answer the research question.

Most authors begin with a literature search. PubMed, Embase, MEDLINE and Google Scholar are databases for scientific articles where most medical published studies can be found. In addition, authors can reach out to subject matter experts (SME) for their thoughts on the topic. Although difficult to cite, SMEs have an understanding of current evidence and the latest ongoing research and can provide feedback regarding the planned search and direct one towards appropriate resources.

One often overlooked area for understanding emerging topics of interest in a given field is through the 'grey literature'. This includes conference proceedings, reports and other documents that are not published in scientific journals. In most cases these sources are not peer-reviewed, and thus their level of quality can be quite varied. Conference proceedings can be reviewed prior to acceptance, but it can be based on incomplete data, as the authors can choose to present preliminary findings before the study is complete [12].

36.4 How to Get Started

36.4.1 Background Research

The first step before choosing both a research topic/question and type of review to address it is to get a preliminary sense of the available literature. Background research can inform how best to identify more sources of information and what search terms may be relevant. The amount and quality of literature can also help determine whether one should write a narrative review or systematic review.

36.4.2 The Outline

After deciding on an appropriate topic, the next step is to design an outline for the review. There is no set structure for designing a narrative review, but typically narrative reviews of medical literature include **introduction** and **discussion** sections [7]. In contrast, systematic reviews have a well-established structure. They require a con-

densed abstract, an introduction, a reproducible description of the methodology, a summary of the available literature in the results and a discussion section drawing overall conclusions from the findings. Guidelines, such as the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) or the Consolidated Standards of Reporting Trials (CONSORT), provide reporting standards that many journals now require be implemented in a published systematic review to aid in their critical appraisal and interpretation [14, 15, 18]. More recently, journals have encouraged the registration of systematic reviews. The Cochrane Library (<http://cochranelibrary.com>) and PROSPERO (<http://crd.york.ac.uk/prospéro/>) offer repositories for ongoing and completed reviews.

Regardless of how the review is structured, the outline should be as detailed as possible. It should address the planned structure of the paper and what information will be addressed in each section. Creating an outline early on allows the paper to be revised without interruption to the literary flow and to coordinate what key messages are to be conveyed throughout each section. Ideally, when using this outline to write the article, the author should be able to complete the paper without having to conduct additional research. More details about conducting a systematic review can be found in the next chapter.

36.5 Writing the Review

With the outline completed, the next step is to write the review. While this may seem like a difficult task at first, there are strategies that can make the job easier.

The first is the use of reference management software that can be found online [e.g. Mendeley (Elsevier, 2017), BibMe (Chegg, 2017)]. Systematic reviews in particular often require an extensive number of articles to be screened and referenced in the final paper. The software can act as an organized repository of information during manuscript preparation. The majority of reference management software can also automate formatting of references to save time when it comes time to submit the review to a journal.

Another strategy is the order in which the review is written. The important sections of a narrative review of medical literature include the figures and tables (when appropriate), the abstract, the title and the main text. The main text of literature reviews can be further broken down into the introduction and discussion sections, with some authors choosing to include methods and/or results sections. In comparison, systematic reviews are required to report all of these components in order to remain transparent.

Figures and tables can also be drafted early on to help organize the structure of the review and focus the main findings. It is essential to create these early in the writing process, as they will be referred to throughout the review. Systematic reviews typically include a flow diagram depicting the screening process (following PRISMA guidelines), a study characteristics table and the detailed search terms in an appendix to ensure the results can be reproduced. Narrative reviews may or may not include these figures or tables.

36.6 Reviews in the Orthopaedic Literature: A Cautionary Tale

Narrative reviews are not typically relied upon within orthopaedics to guide clinical decisions. This weight falls upon systematic reviews, which are historically cited more than their nonsystematic counterparts [2]. In the past, surgical literature, including orthopaedic literature, has been found to be lacking in quality [4]. Bhandari et al. applied the Detsky scale to assess the reporting quality of 72 randomized controlled trials (RCTs) published in the *Journal of Bone and Joint Surgery* from 1988 to 2000. Only 32 (43%) of those studies were found to be of high quality [3]. The poor quality of this literature was attributed to two reasons. The first was a reliance by systematic reviews on low-quality evidence, such as case studies and case series. These are inherently limited by their retrospective nature, potential for bias and lack of comparative groups. The second is that high quality systematic reviews regularly failed to sufficiently report the study parameters, including methodological parameters of the study, their sources of funding and of potential

conflicts, as well as the quality of their evidence [4]. In the last decade, there has been a dramatic growth in the number of randomized controlled trials and systematic reviews published in orthopaedic literature. Publications in orthopaedics on the whole are estimated to have doubled between 2000 and 2011 [9]. However, the quality of these publications and reviews were still poor, despite publishing in top journals [6]. This reduces the impact of published systematic reviews and their ability to aid in clinical decision making. By using the strategies described in this chapter to carefully plan any type of review, authors can minimize bias and ensure adequate reporting to demonstrate higher quality.

36.7 Conclusion

Narrative reviews can be written relatively quickly and often provide the most current and up-to-date information on a chosen topic. They are commonly written by authors who are experts in the discussed topic, allowing those experts to put forward their educated opinions and ideas. Systematic reviews are conducted using a planned methodological structure, they attempt to encompass all available literature for a chosen topic, and they provide reproducible and typically less biased results. This chapter has elucidated the differences between these two types of reviews and serves to provide a starting point for any author thinking of conducting a narrative review of their own.

References

- Alper B, Hand J, Elliott S. How much effort is needed to keep up with the literature relevant for primary care. *J Med Libr Assoc.* 2004;92:429.
- Bhandari M, Montori VM, Devereaux PJ, Wilczynski NL, Morgan D, Haynes RB, Hedges Team. Doubling the impact: publication of systematic review articles in orthopaedic journals. *J Bone Joint Surg Am.* 2004;86:1012–6.
- Bhandari M, Richards RR, Sprague S, Schemitsch EH. The quality of reporting of randomized trials in the *Journal of Bone and Joint Surgery* from 1988 through 2000. *J Bone Joint Surg Am.* 2002;84:388–96.
- Chaudhry H, Mundi R, Singh I, Einhorn TA, Bhandari M. How good is the orthopaedic literature? *Indian J Orthop.* 2008;42:144–9.
- Ethgen O, Bruyère O, Richey F, Dardennes C, Reginster J-Y. Health-related quality of life in total hip and total knee arthroplasty. A qualitative and systematic review of the literature. *J Bone Joint Surg Am.* 2004;86:963–74.
- Gagnier JJ, Kellam PJ. Reporting and methodological quality of systematic reviews in the orthopaedic literature. *J Bone Joint Surg Am.* 2013;95:e771–7.
- Green B, Johnson C, Adams A. Writing narrative literature reviews for peer-reviewed journals: secrets of the trade. *J Chiropr Med.* 2006;5:101–17.
- Hitchcock M. Review vs systematic review vs ETC. In: *LibGuides Nurs. Resour.* 2017. <http://researchguides.ebling.library.wisc.edu/c.php?g=293229&p=1953452>. Accessed 6 Oct 2017.
- Hui Z, Yi Z, Peng J. Bibliometric analysis of the orthopedic literature. *Orthopedics.* 2013;36:e1225–32.
- Hurwitz S, Slawson D, Shaunessy A. Orthopaedic information mastery: applying evidence-based information tools to improve patient outcomes while saving orthopaedists' time. *J Bone Joint Surg Am.* 2000;82(6):888–94.
- Hussain N, Turvey S, Bhandari M. Keeping up with best evidence: what resources are available? *J Postgrad Med Edu Res.* 2012;46:4–7.
- Kay J, Memon M, Rogozinsky J, de Sa D, Simunovic N, Seil R, Karlsson J, Ayeni OR. The rate of publication of free papers at the 2008 and 2010 European Society of Sports Traumatology Knee Surgery and Arthroscopy congresses. *J Exp Orthop.* 2017;4:15.
- Miller LE, Gondusky JS, Bhattacharyya S, Kamath AF, Boettner F, Wright J. Does surgical approach affect outcomes in total hip arthroplasty through 90 days of follow-up? A systematic review with meta-analysis. *J Arthroplasty.* 2017;33(4):1296–302.
- Moher D. Consort: an evolving tool to help improve the quality of reports of randomized controlled trials. *JAMA.* 1998;279:1489–91.
- Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev.* 2015;4:1.
- Riemsma RP, Pattenden J, Bridle C, Sowden AJ, Mather L, Watt IS, Walker A. Systematic review of the effectiveness of stage based interventions to promote smoking cessation. *Br Med J.* 2003;326:1175–7.
- Rother ET. Systematic literature review X narrative review. *Acta Paul Enferm.* 2007;20:5–6.
- Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMC Med.* 2010;8:18.
- Young NL, Cheah D, Waddell JP, Wright JG. Patient characteristics that affect the outcome of total hip arthroplasty: a review. *Can J Surg.* 1998;41:188–95.

Part VII

How to Perform a Systematic Review or Meta-analysis?



What Is the Difference Between a Systematic Review and a Meta-analysis?

37

Shakib Akhter, Thierry Pauyo, and Moin Khan

37.1 Hierarchy of Evidence

The hierarchy of evidence serves as the foundation for evidence-based practice and provides a top-down descriptive visualization of the best available evidence. The level of evidence is proportional to reliability, quality, and validity; the higher these factors, the higher the study lies in the hierarchy of evidence [2, 27]. Clinicians and scientists seek higher levels of evidence as these studies have the greatest potential impact for clinical practice [3]. Experimental study designs, such as randomized controlled trials, occupy the highest level of evidence (level 1 evidence) in research, followed by observational study designs, such as cohort studies (level 2 evidence) [2, 3, 27]. These research designs are followed by case control studies (level 3 evidence), case studies and case series (level 4 evidence), and finally expert opinions (level 5 evidence) [2, 3, 27]. Clinicians should cautiously apply low-quality evidence such as the latter designs given their poor reliability, reproducibility, validity, and clin-

ical impact [2, 3, 27]. Higher-quality research evidence (such as level 1 and level 2) is readily brought from bench to bedside as their superior methods increase study validity. But the question arises: where do systematic reviews and meta-analysis fall in this hierarchy of evidence? The answer is that they each fall in every level. A systematic review and meta-analysis of well-done high-quality randomized controlled trials are considered the pinnacle of evidence-based research. It is important to understand and appreciate that systematic reviews and meta-analyses are not always atop the hierarchy of evidence; if a researcher conducts a systematic review of level 3 evidence, the review will be considered level 3 evidence. Conversely, a systematic review of level 1 evidence will remain level 1 evidence. Accordingly, the study design and level of evidence are directly proportional. Systematic reviews and meta-analyses remain confined to the level of evidence they are used with but provide invaluable results and have tremendous clinical care implications.

S. Akhter (✉) · M. Khan
Department of Orthopaedic Surgery,
McMaster University, Hamilton, ON, Canada

Department of Health, Evidence, and Impact,
McMaster University, Hamilton, ON, Canada
e-mail: akhterms@mcmaster.ca

T. Pauyo
Department of Orthopaedic Surgery,
McGill University, Montreal, QC, Canada

37.2 Why Perform a Systematic Review or Meta-analysis?

With over 50 million scholarly articles published to date, difficulty exists among clinicians and scientists in organizing and understanding the vast amounts of available literature [10]. Arguably,

the most effective and efficient method in synthesizing available data in order to make evidence-based decisions is by conducting systematic reviews and meta-analyses. These methods identify, critically appraise, and evaluate several studies pertaining to a single prespecified research question [21]. Both methodologies allow researchers to combine large portions of literature and produce results that are widely generalizable. Additional benefits include the ability to limit biases found in individual studies, which increases the reliability of the findings. Although the systematic review and meta-analysis methodologies synergistically analyse numerous studies, they both serve different yet complementary roles. Systematic reviews are the cornerstone in evidence-based medicine as the meticulous, exhaustive, systematic, and structured approach is effective in summarizing and critically analysing the findings of relevant literature pertinent to a research questions. A meta-analysis may be conducted in conjunction with a systematic review to enhance the credence of findings by increasing the level of evidence. Through combining and analysing several studies, researchers aim to extrapolate results that more accurately reflect true effects. Synthesizing data from multiple studies allows researchers to achieve a greater level of statistical power, which is precluded in individual studies. Researchers practising this methodology are encouraged to follow criteria outlined by research quality improvement bodies, such as the Preferred Reporting Items for Systematic Reviews and Meta-analysis (PRISMA) statement [16].

37.3 Systematic Review

37.3.1 What Is a Systematic Review?

Literature reviews are commonly classified as either a narrative or systematic review. A narrative review is a synthesis of literature in a specific field constructed using a specific contextual or theoretical point of view. The method is not nearly as robust or rigorous as no methodological approaches, such as the search strategy or

eligibility criteria, are reported. Narrative reviews employ a qualitative approach to critically appraise literature from a specific context or theoretical perspective. A narrative review is highly susceptible to bias and lacks reproducibility. Researchers supplement narrative reviews with expert intuitive and experiential evidence, reflecting the qualitative approach [20]. The lack of a systematic process predisposes this type of review to bias, specifically subjective selection bias, which has implications on the validity and generalizability of the study.

A systematic review is governed by carefully constructed stages which guide the cumulative search method, screening, reviewing, cataloguing, and reporting process of the selected studies to answer a research question of interest. The search is intended to provide an exhaustive review of the current literature, which captures all appropriate studies ensuring the research question to be answered to its fullest extent. The carefully crafted and exhaustive methodology minimizes bias and provides reliable and accurate conclusions to be drawn. It has facilitated an improved dissemination of information to healthcare providers, the public, policymakers, and the scientific community. A notable benefit is the resulting mitigation in the delay in bringing research from bench to bedside leading to well-informed policy decisions and direction for future research. Although systematic reviews rank high in the hierarchical chain of evidence, they are not without flaws. Depending on the construction of the search and screening criteria, a researcher may inadvertently exclude relevant studies. However, given the systematic, explicit, and comprehensive methodology of the review, the risk of missing relevant studies is decreased in comparison to a narrative review. Additionally, the generalizability of the findings may be limited as patients included in the review must be homogenous to the target population the researcher attempts to generalize. Heterogeneity is a marked concern among researchers as a balance between internal and external validity must be prioritized; although stringent eligibility criteria produce homogenous data, generalizability is affected as patients with differing characteristics

may be excluded [28]. The systematic review methodology is indicated in this context as it can provide a descriptive and analytical summary of heterogeneous and dissimilar studies, which usually preclude the use of a meta-analysis [28].

37.3.2 A Guide to Performing a Systematic Review

Six essential steps in conducting a systematic review are described.

Step 1: The Research Question—The first and foremost step in the process of performing quality research is formulating appropriate research questions. Commonly underestimated, formulating the research question requires careful consideration of a multitude of factors and can potentially consume a notable amount of time. The research question should include the problem or question of interest, population of interest, intervention, and comparator, along with the outcomes of interest [1]. Using the population, intervention, comparator, and outcomes (PICO) format helps ensure the question is direct and relevant. Although sometimes loosely stated, authors should explicitly present the question in a structured format as it allows consumers to directly understand the goal of the project [12].

Step 2: Constructing Eligibility Criteria—Selecting appropriate eligibility criteria is fundamental in ensuring the most relevant and appropriate studies are included in the review. Researchers can include all key components needed to comprehensively answer their question if the criteria are reflective of the PICO points identified in Step 1. Additionally, researchers should clearly identify which type of studies they are interested in (i.e. randomized trials, cohort studies, etc.) and other operational factors relevant to the studies including, but not limited to, year published, language, and number of participants [25].

Step 3: Constructing the Search Strategy—Using appropriate search terms is essential in ensuring an exhaustive search of literature relevant to the research question. An incomplete search can exclude important studies that

may have significant implications to the validity of the review. An extensive search without language restrictions is recommended [12]. A comprehensive search often has three or more online databases included. Examples of commonly used databases include MEDLINE, EMBASE, CENTRAL, CINAHL, PUBMED, and others. These platforms allow the researcher to conduct a comprehensive and exhaustive search of literature published on the World Wide Web, download results, and perform a citation analysis [17]. A critical tool for success, although not necessary, but highly recommended, is seeking a professional librarian to assist in the search strategy development and implementation [25]. The terms should also be reflective of the PICO format to produce an effective search strategy that identifies all relevant studies and balances sensitivity and specificity [25]. In this context, good sensitivity and specificity are reached when the search produces a high number of relevant studies and a low number of irrelevant studies, respectively. With an increasing trend in technological dominance in research, many authors chose to solely search electronic databases, but searches of reference lists on articles and specific journals can also be done [25]. A search of not only published data via online databases but also of ‘grey literature’ decreases the risk for potential publication bias of results. Grey literature refers to scholarly literature that is not formally published and can include dissertations, policy documents, book chapters, and conference abstracts, research reports, and unpublished research data [9].

Step 4: Screening, Selection, and Extraction—This process begins upon completion of the search. A list of all article abstracts that appeared in the search undergoes an abstract screening, in which researchers make decisions about including or excluding studies based on their relevance to the eligibility criteria elicited from the abstract. A recommended practice to establish inter-rater reliability and increase validity of the study is to have two independent reviewers conduct this process independently. Often, reviewers may face disagreements whether to include or exclude studies. Attempts to reach agreements between the two reviewers should be made, but

if unsuccessful, another study researcher should be consulted. Researchers should log all activities in this process, such as which studies were included and excluded and the reasons for each [25]. This log will help researchers create a flow diagram to visually represent articles included. Figure 37.1 is an example taken from the full-text article of Clinical Vignette 2. Once all titles and abstracts are screened and a significantly shorter list of studies remains, a second full-text screening process should be conducted by the same reviewers to ensure reliability and consistency in study methods. This process will exclude all remaining irrelevant studies and identify which have extremely high potential to be included in the review. Finally, researchers should create a data table or a standardized data extraction form to systematically organize and extract data. This table or form is unique to each systematic review and may include patient clinical and demographic characteristics, author names, type of study design, and outcomes. It is most commonly electronic (i.e. a constructed table in excel) given its associated considerable efficiency and mitigation of data errors due to mismanagement, but paper also may be used

[28]. The data extraction process also serves as the final screening process, as the extracted data will guide final decisions regarding which studies will be included or excluded [25, 28]. Researchers must also be prepared for when they face studies that have missing or incomplete data, for which they must contact the individual study authors to obtain [16].

Step 5: Quality Assessment—Appraising the study quality, although not required, is highly recommended to produce a comprehensive and highly valid systematic review. A standardized operational definition of ‘study quality’ is elusive but generally refers to the confidence a researcher holds in the study’s design and methods to minimizing bias [19, 21, 28]. In other words, a quality assessment is a critical appraisal to determine if the design, conduct, and methods of a study will reduce systematic error and bias, which in turn has implications on its internal and external validity [4, 28]. Bias is defined by the Cochrane Collaboration as a deviation from truth or a systematic error that can lead to either an under- or overestimation of the true effect [6]. There are multiple types of bias (selection, detection, reporting, attrition, performance biases) that range from small to substantial and can markedly limit the generalizability of a research study [6]. However, researchers should be cautious in the interpretation of a quality assessment as the process is inherently limited by factors such as a lack of information provided by the author as well as the absence of a ‘gold standard’ to reference level of quality [21, 26]. Nonetheless, recommended guidelines such as the five-point Oxford Quality Rating Scale or the Consolidated Standards of Reporting Trials (CONSORT statement) are available for comprehensive quality assessments and should be used when appropriate. It is also recommended that at least two independent reviewers conduct the quality assessment process and increase inter-rater reliability and study validity [8]. A consensus within available literature outlines four main biases that affect study quality: performance, selection, attrition, and detection bias [4, 13, 24, 28]. The quality assessment is a critical step in the systematic review process as these

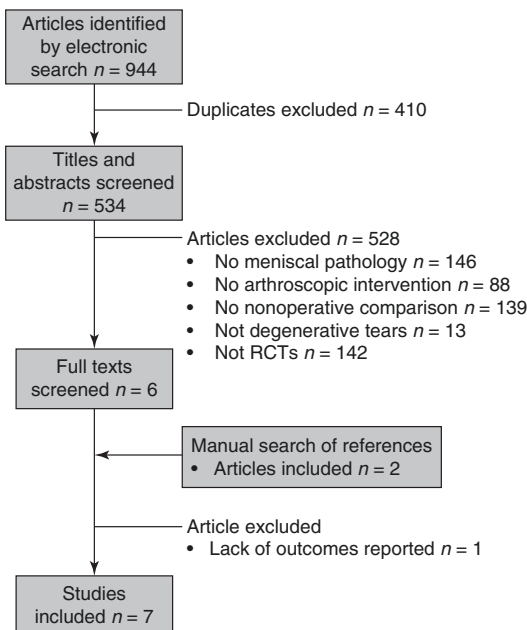


Fig. 37.1 Article selection flow diagram

biases can lead to under- or overestimations of the true effect, and the resulting spurious associations may have detrimental implications for clinical practice [12, 28].

Quality assessment tools depend on what type of evidence is being synthesized. The Cochrane risk of bias (ROB) tool is considered the gold standard for the quality assessment of randomized controlled trials, whereas the methodological index for non-randomized studies (MINORS) is one of many available measures for assessing observational studies. The Cochrane ROB tool ensures a systematic and critical appraisal of all possible bias domains including selection, performance, detection, attrition, reporting, and other forms of biases [29]. Assessing for these potential biases ensures the researcher takes into account systematic differences in baseline characteristics, provision of care, outcome ascertainment, participant withdrawals, and reported and unreported findings within the studies. The MINORS is a 12-item tool that is used for, but not limited to, quality assessment of systematic reviews that employ observational studies (Fig. 37.2) [23]. This tool has demonstrated

strong external validity through inter-reviewer agreement, reliability via the kappa coefficient, and consistency via Cronbach’s alpha coefficient [23]. The kappa coefficient is a commonly used statistical measure of inter-rater reliability, and Cronbach’s alpha coefficient elicits the consistency and correlation of the items of an instrument, such as a survey, to determine its overall reliability [18, 22].

Providing an overall confidence in the estimate of effect of a particular treatment or intervention is increasingly being performed utilizing the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) criteria. This GRADE method provides a reliable estimate of the effect of the evidence across all studies for outcomes, as opposed to individual study evaluation. An example of one of the factors critiqued by GRADE is shown in Fig. 37.3. Methodological flaws, treatment effects, consistency, and generalizability comprise the main assessment domains seen in this tool [5]. Great care should be taken to select the appropriate tool for quality assessment and should be reflective of the type of studies being reviewed.

Methodological items for non-randomized studies	Score†
<ol style="list-style-type: none"> 1. A clearly stated aim: the question addressed should be precise and relevant in the light of available literature 2. Inclusion of consecutive patients: all patients potentially fit for inclusion (satisfying the criteria for inclusion) have been included in the study during the study period (no exclusion or details about the reasons for exclusion) 3. Prospective collection of data: data were collected according to a protocol established before the beginning of the study 4. Endpoints appropriate to the aim of the study: unambiguous explanation of the criteria used to evaluate the main outcome which should be in accordance with the question addressed by the study. Also, the endpoints should be assessed on an intention-to-treat basis. 5. Unbiased assessment of the study endpoint: blind evaluation of objective endpoints and double-blind evaluation of subjective endpoints. Otherwise the reasons for not blinding should be stated 6. Follow-up period appropriate to the aim of the study: the follow-up should be sufficiently long to allow the assessment of the main endpoint and possible adverse events 7. Loss to follow up less than 5%: all patients should be included in the follow up. Otherwise, the proportion lost to follow up should not exceed the proportion experiencing the major endpoint 8. Prospective calculation of the study size: information of the size of detectable difference of interest with a calculation of 95% confidence interval, according to the expected incidence of the outcome event, and information about the level for statistical significance and estimates of power when comparing the outcomes <p><i>Additional criteria in the case of comparative study</i></p> <ol style="list-style-type: none"> 9. An adequate control group: having a gold standard diagnostic test or therapeutic intervention recognized as the optimal intervention according to the available published data 10. Contemporary groups: control and studied group should be managed during the same time period (no historical comparison) 11. Baseline equivalence of groups: the groups should be similar regarding the criteria other than the studied endpoints. Absence of confounding factors that could bias the interpretation of the results 12. Adequate statistical analyses: whether the statistics were in accordance with the type of study with calculation of confidence intervals or relative risk 	

Each domain is scored from 0-2. Items will either be scored as 0 (not reported), 1 (reported but inadequate), or 2 (reported and adequate) [29]. A score for non-comparative studies of 16 and comparative studies of 24 is globally accepted [29].

Fig. 37.2 The validated and updated version of the MINORS questionnaire [29]

Risk of bias	Across studies	Interpretation	Considerations	GRADE assessment of study limitations
Low risk of bias.	Most information is from studies at low risk of bias.	Plausible bias unlikely to seriously alter the results.	No apparent limitations.	No serious limitations, do not downgrade.
Unclear risk of bias.	Most information is from studies at low or unclear risk of bias.	Plausible bias that raised some doubt about the results.	Potential limitations are unlikely to lower confidence in the estimate of effect.	No serious limitations, do not downgrade.
			Potential limitations are likely to lower confidence in the estimate of effect.	Serious limitations. downgrade one level.
High risk of bias.	The proportion of information from studies at high risk of bias is sufficient to affect the interpretation of results.	Plausible bias that seriously weakens confidence in the results.	Crucial limitation for one criterion, or some limitations for multiple criteria, sufficient to lower confidence in the estimate of effect.	Serious limitations. downgrade one level.
			Crucial limitation for one or more criteria sufficient to substantially lower confidence in the estimate of effect.	Very serious limitations, downgrade two levels.

Adopted from the Cochrane Handbook of Systematic Reviews of Interventions, this is an extension of factor 1 of 5 of GRADE that extends the risk of bias assessment to make conclusions regarding the limitations relating to outcomes [6].

Fig. 37.3 Factor 1 of 5 in GRADE assessment [6]

A comprehensive quality assessment adds to the review’s generalizability and validity and therefore leads to a greater clinical impact.

Step 6: Data Analysis and Interpretation of Results—The final steps of data analysis and interpretation should be conducted following a comprehensive quality assessment [4, 13, 24, 28]. This step requires the researcher to create a concise and simple descriptive summary of each included study [28]. Many researchers choose to display study characteristics in a tabular format. An example can be found in Table 1 in the full article of the Clinical Vignette 1. The descriptive table is a comprehensive summary of study characteristics and therefore should include, but is not limited to, author name, year published, study

methods, biases, quality assessment, total number of patients in treatment and control groups, intervention(s), control, outcomes, and any other relevant information. The inclusion of items in the table should reflect the research question [20]. Interestingly, these information-rich tables provide sufficient information for the researcher to determine if statistically pooling the data for a meta-analysis is possible [20]. If the study includes sufficient data to be meta-analysed, that approach will be adopted as it is a higher level of evidence. As a result, a systematic review can be, and is commonly, accompanied by a meta-analysis (Clinical Vignette 2).

When interpreting results, the researcher should carefully extrapolate conclusions based

on the summarization and analysis of the included studies. In this section, the findings must be briefly reiterated, and the final impressions from this synthesis of the best available evidence should be explicitly stated. Here, the researcher should highlight the strengths and limitations of their study and indicate how these findings may or may not have clinical implications. Suggestions for future directions of research should also be included as they are valuable statements in systematic reviews, provided this methodology summarizes available evidence pertaining to a specific field and highlights lacking areas. Systematic reviews directly improve patient care as they provide information that is a requisite for evidence-based decision-making.

Clinical Vignette 1 displays a systematic review on the efficacy of antifibrinolytic therapy in reducing patient transfusions in orthopaedic surgery. The authors clearly and explicitly state

their research question of interest in the background section (Step 1). The eligibility criteria and the use of two independent reviewers were highlighted (Step 2). Four databases were searched producing 283 articles, of which 29 were included following the screening process (Step 4). Results and figures for study and patient characteristics, along with results for study quality and clinical outcomes, were reported (Step 5). This information was also visually displayed using histograms. The results in each individual studies were reported cumulatively as a single article of evidence, without the use of statistical methods. The authors finally critically interpreted the results in the discussion section (Step 6). Evidently, the authors clearly report their methods allowing reproducibility and demonstrating a strong internal validity in their review.

Clinical Vignette 1: Systematic Review [11]

Review Article

Use of antifibrinolytic therapy to reduce transfusion in patients undergoing orthopedic surgery: A systematic review of randomized trials

Yoan K. Kagoma^a, Mark A. Crowther^{b,*}, James Douketis^b, Mohit Bhandari^c, John Eikelboom^b, Wendy Lim^b

^a Applied Science and Engineering, University of Toronto, Toronto, Ontario, Canada

^b Departments of Medicine and Hematology, McMaster University (St. Joseph's Hospital / Hamilton General Hospital), Hamilton, Ontario, Canada

^c Department of Clinical Epidemiology and Biostatistics and Surgery (Division of Orthopedics), McMaster University, Hamilton, Ontario, Canada

ARTICLE INFO

Article history:

Received 26 May 2008

Received in revised form 4 September 2008

Accepted 29 September 2008

Available online 12 November 2008

ABSTRACT

Background: Minimizing bleeding and transfusion is desirable given its cost, complexity and potential for adverse events. Concerns have been heightened by recent data demonstrating that bleeding events may predict worse outcomes and by warnings about the safety of erythropoietic stimulating agents. Prior small studies suggest that antifibrinolytic agents may reduce bleeding and transfusion need in patients undergoing total hip replacement (THR) or total knee arthroplasty (TKA). However, no single study has been large enough to definitively determine if these agents are safe and effective. To address this issue we performed a systematic review of randomized trials describing the use of tranexamic acid, epsilon aminocaproic acid, or aprotinin administration in the perioperative setting.

Methods: MEDLINE, EMBASE, CINAHL and the Cochrane databases were searched for relevant trials. Two independent reviewers abstracted total blood loss, transfusion requirements, and venous thromboembolism (VTE) rates. Data were combined using the Mantel-Haenszel method and dichotomous data expressed as relative risk (RR) with 95% confidence intervals (CI).

Results: Patients receiving antifibrinolytic agents had reduced transfusion need (RR 0.52; 95% CI, 0.42 to 0.64; $P < 0.00001$), reduced blood loss and no increase in the risk of VTE (RR 0.95% CI, 0.80 to 1.10, $I^2 = 0\%$, $P = 0.531$).

Conclusions: We conclude that antifibrinolytic agents may reduce bleeding and transfusion in patients undergoing THR or TKA who receive appropriate antithrombotic prophylaxis. There is a need for a large, adequately powered prospective study to carefully examine the safety and efficacy of these agents.

© 2008 Published by Elsevier Ltd.

Fact Check

- A systematic review is a comprehensive and exhaustive review of relevant literature specific to a topic or research question.
- Systematic reviews may employ qualitative or quantitative methods in reviewing and reporting homogenous or heterogeneous studies.
- Compared to the meta-analysis method, a systematic review can be completed readily and requires less formal training, but is of a lower level of evidence.
- The six steps in performing a comprehensive systematic review include (1) formulating an appropriate research question; (2) formulating appropriate eligibility criteria; (3) forming and conducting an appropriate search strategy; (4) screening, selection, and extraction of relevant results; (5) quality assessment of the included studies; and (6) critical appraisal to summarize and interpret the findings.

meaningful results, for example, by detecting modest associations. This design also provides researchers a comprehensive understanding of the true effect. Clinical, methodological, and statistical heterogeneities are issues of variability that impose limitations on the systematic review and meta-analysis methodologies. With limited heterogeneity, more data provides a more precise estimate of effect. Moreover, the utilization of statistical methods addresses notable concerns of generalizability and bias, which are inherent in the systematic review alone. With respect to heterogeneity, the meta-analysis quantifies between group differences (differences between studies) and also explains them (e.g. using tools such as a meta-regression). However, the presence of excessive heterogeneity may result in faulty and misleading conclusions. Statistical methods to test for heterogeneity are discussed below. Although this methodology occupies the top of the hierarchy of evidence, as every research design, it has limitations [7]. The nature of the meta-analysis requires a large number of studies in order for an effect to be seen. Additionally, although it can combine the data of various studies, the meta-analysis is not capable of adjusting for poor methodology of the studies included which may skew the outcomes. It is important for the researcher to create specific eligibility criteria, so that the literature search and review process is maximally exhaustive and produces methodologically sound studies.

37.4 Meta-analysis

37.4.1 What Is a Meta-analysis?

A meta-analysis (Clinical Vignette 2), much like a systematic review and often an extension of one, also hinges on a systematic and exhaustive search of the literature. A meta-analysis differs from a systematic review in that instead of simply collecting and analysing the data, it employs statistical methods to quantitatively synthesize the results from multiple studies [15, 28]. This design aims to expose true effects buried within data by analysing patterns to compare and contrast findings of several studies. A meta-analysis has many inherent benefits compared with a systematic review. A notable advantage is that pooling studies increases statistical power, otherwise unattainable in individual studies, leading to more

37.4.2 A Guide to Performing a Meta-analysis

The six steps in performing a meta-analysis are the same in performing a systematic review. The only difference is additional methods are utilized to perform a meta-analysis in the data analysis in Step 6. The steps follow this chronological order: (1) formulating an appropriate research question; (2) formulating appropriate eligibility criteria; (3) forming and conducting an appropriate search strategy; (4) screening, selection, and extraction of relevant results; (5) quality assessment of the

included studies; and (6) analysing results and interpreting the findings.

Step 6: Data Analysis and Interpretation of Results—Calculation of the effect sizes and reporting them with a 95% confidence interval (CI) are common practice with meta-analyses [25]. Numerous statistical programmes including the Cochrane-endorsed Review Manager programme (RevMan), MIX 2.0, and MetaStat that conduct the meta-analysis processes are widely available. The results (effect sizes and CI intervals) should be reported graphically and quantitatively [25]. A common graphical representation of meta-analysis results is a forest plot (Fig. 37.4). In Fig. 37.4, the forest plot visually depicts each study as a square where the middle is the effect size (SMD) and each end point of the corresponding line represents the upper and lower CI limits. The right portion of the plot (>0) favours the control or comparator, whereas the left (<0) favours the intervention [25]. The large diamond at the bottom represents the pooled effect of all the individual studies. As the left side of the graph favours the intervention, researchers hope to see this diamond, or pooled effect, below <0 indicating efficacy of the intervention. Calculation of inter-rater agreement and tests of heterogeneity are also done in this phase. Measurement of inter-rater reliability is a key component in the validity of a study as it represents how well the data collected are accurate representations of the variables of interest by quantitatively measuring the extent to which

raters (data collectors) agree in their independent measurements [18]. A multitude of statistical methods to test for inter-rater reliability exist, including percent agreement, the contingency coefficient, Pearson’s r, the correlation coefficient, the concordance correlation coefficient, and the most commonly used Cohen’s kappa for two raters or Fleiss kappa for three or more raters [18]. Although heterogeneity can be judged from graphical representations of data, such as looking at the error bars in forest plots, researchers should conduct a statistical test of heterogeneity to address concerns of dissimilarities in study results within the meta-analysis. This test determines if the variation in the study results is due to genuine measurable differences (heterogeneity) or chance alone (homogeneity) but has the inherent limitation of sensitivity to the number of included trials [8]. The *I*² statistic is commonly used to quantitatively measure the variability between results (effect sizes of each study). This assessment of consistency is critical as it directly relates to generalizability; the more consistency in studies, the more generalizable it is [18]. The statistic Cochran’s *Q* is commonly used to evaluate the null hypothesis of all included studies and evaluate similar effects [8]. This test statistic is calculated by weighing the sums of squared deviated from the individual studies and the overall pooled result [14]. Finally, a chi-squared (χ^2) distribution with *k*-1 degrees of freedom is compared to the results from the *Q* test statistic to obtain *P* values [9].

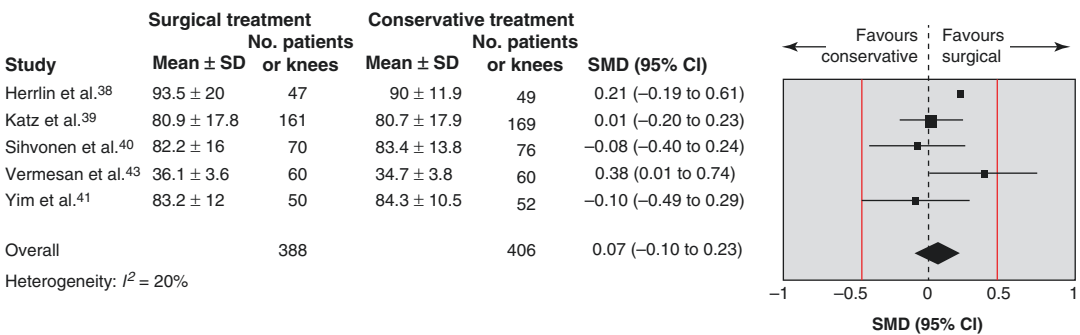


Fig. 37.4 Meta-analysis forest plot

With interpreting the results, the researcher should practise similar caution outlined in the systematic review process. Summarizing findings, making evidence-based conclusions, highlighting patient care implications, and suggesting future directions for research should be included.

Clinical Vignette 2 [14] displays a meta-analysis on arthroscopic surgery for degenerative tears of the meniscus. Authors systematically searched three databases to ensure all relevant literature is captured. With a focus on randomized controlled trials, the search was conducted and studies screened and assessed for eligibility by two independent reviewers. To ensure a systematic and consistent process, the same reviewers independently conducted risk of bias assessments followed by data extraction using

an electronic extraction form. The authors employed statistical methods to report interobserver agreement and outcomes. All steps of the statistical methods are expertly reported and can be readily reproduced. Test of heterogeneity was also performed to determine if variability was a result of chance or inter-study heterogeneity. Subgroup analyses were outlined a priori, and sensitivity analyses were done to elucidate the consequences of missing data and studies at risk for bias. Figures of the study selection process and study descriptions were provided. Results for the search, each individual outcome, adverse events, and sensitivity analysis were reported and interpreted. Authors followed the PRISMA statement for reporting findings. The authors concluded with limitations, implications, and a conclusion.

Clinical Vignette 2: Meta-analysis

Arthroscopic surgery for degenerative tears of the meniscus: a systematic review and meta-analysis

Moin Khan MD, Nathan Evaniew MD, Asheesh Bedi MD, Olufemi R. Ayeni MD MSc, Mohit Bhandari MD PhD

ABSTRACT

Background: Arthroscopic surgery for degenerative meniscal tears is a commonly performed procedure, yet the role of conservative treatment for these patients is unclear. This systematic review and meta-analysis evaluates the efficacy of arthroscopic meniscal débridement in patients with knee pain in the setting of mild or no concurrent osteoarthritis of the knee in comparison with nonoperative or sham treatments.

Methods: We searched MEDLINE, Embase and the Cochrane databases for randomized controlled trials (RCTs) published from 1946 to Jan. 20, 2014. Two reviewers independently screened all titles and abstracts for eligibility. We assessed risk of bias for all included studies and pooled outcomes using a random-effects model. Outcomes (i.e., function and pain relief) were dichotomized to short-term (< 6 mo) and long-term (< 2 yr) data.

Results: Seven RCTs ($n = 805$ patients) were included in this review. The pooled treatment

effect of arthroscopic surgery did not show a significant or minimally important difference (MID) between treatment arms for long-term functional outcomes (standardized mean difference [SMD] 0.07, 95% confidence interval [CI] -0.10 to 0.23). Short-term functional outcomes between groups were significant but did not exceed the threshold for MID (SMD 0.25, 95% CI 0.02 to 0.48). Arthroscopic surgery did not result in a significant improvement in pain scores in the short term (mean difference [MD] 0.20 , 95% CI -0.67 to 0.26) or in the long term (MD -0.06 , 95% CI -0.28 to 0.15). Statistical heterogeneity was low to moderate for the outcomes.

Interpretation: There is moderate evidence to suggest that there is no benefit to arthroscopic meniscal débridement for degenerative meniscal tears in comparison with nonoperative or sham treatments in middle-aged patients with mild or no concomitant osteoarthritis. A trial of nonoperative management should be the first-line treatment for such patients.

Competing interests:

Mohit Bhandari declares consultancy payments from Smith & Nephew, Stryker, Amgen, Zimmer, Moximed and Bioventus, and grant support from Smith & Nephew, DePuy, Eli Lilly and Bioventus. No other competing interests were declared.

This article has been peer reviewed.

Correspondence to:

Moin Khan, moinkhanmd@gmail.com

CMAJ 2014. DOI:10.1503/cmaj.140433

Fact Check

- A meta-analysis involves a comprehensive and exhaustive review of relevant literature specific to a topic or research question and can be viewed as an extension to a systematic review.
- Meta-analyses employ statistical methods to quantitatively synthesize the results of pooled studies.
- Performing a meta-analysis is more difficult and time-consuming than performing a systematic review but is of a higher level of evidence.
- The six steps in performing a comprehensive systematic review and meta-analysis are the same with one addition. Only step 6 involves an additional data analysis via statistical software such as the Cochrane Collaboration-endorsed ‘RevMan’.

Take-Home Message

- Both the systematic review and meta-analysis are considered the highest level of research evidence.
 - A meta-analysis employs statistical methods to quantitatively synthesize results of different studies and, when pooling high-quality randomized controlled trials, is considered the highest level of evidence.
 - This design recognizes and adjusts results for biases, increasing the internal and external validity of the study.
 - Conversely, a systematic review is a nonstatistical, formal, and structured process governed by carefully constructed stages which guide the cumulative search method, screening, reviewing, cataloguing, and reporting processes of selected studies to answer a research question of interest.
 - Researchers must carefully consider which method is indicated for the question of interest as specific factors preclude specific designs, as seen with heterogeneity precluding the use of a meta-analysis.
- Lastly, researchers should appraise current reviews and meta-analyses for their methods as well as adhere to best practice guidelines to produce research that maximizes both internal and external validity, resulting in clinically relevant implications.

References

1. Aslam S, Emmanuel P. Formulating a researchable question: a critical step for facilitating good clinical research. *Indian J Sex Transm Dis AIDS*. 2010;31(1):47.
2. Brighton B, Bhandari M, Tornetta P, Felson DT. Hierarchy of evidence: from case reports to randomized controlled trials. *Clin Orthop Relat Res*. 2003;413:19–24.
3. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg*. 2011;128(1):305–10.
4. Egger M, Davey-Smith G, Altman D. *Systematic reviews in health care: meta-analysis in context*. Somerset: Wiley; 2013.
5. GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;328(7454):1490.
6. Green S, Higgins JP. Preparing a cochrane review. In: *Cochrane handbook for systematic reviews of interventions*; 2012. p. 11–30.
7. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. *Users’ guides to the medical literature*. IX. A method for grading health care recommendations. Evidence-Based Medicine Working Group. *JAMA*. 1995;274(22):1800–4.
8. Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557–60.
9. Hopewell S, McDonald S, Clarke M, Egger M. Grey literature in meta-analyses of randomized trials of health care interventions. *Cochrane Database Syst Rev*. 2007;(2):MR000010.
10. Jinha AE. Article 50 million: an estimate of the number of scholarly articles in existence. *Learned Publ*. 2010;23(3):258–63.
11. Kagoma YK, Crowther MA, Douketis J, Bhandari M, Eikelboom J, Lim W. Use of antifibrinolytic therapy to reduce transfusion in patients undergoing orthopedic surgery: a systematic review of randomized trials. *Thromb Res*. 2009;123(5):687–96.
12. Khan KS, Kunz R, Kleijnen J, Antes G. Five steps to conducting a systematic review. *J R Soc Med*. 2003;96(3):118–21.
13. Khan KS, Kunz R, Kleijnen J, Antes G. *Systematic reviews to support evidence-based medicine: how to review and apply findings of healthcare research*. London: Royal Society of Medicine Press; 2003.

14. Khan M, Evaniew N, Bedi A, Ayeni OR, Bhandari M. Arthroscopic surgery for degenerative tears of the meniscus: a systematic review and meta-analysis. *Can Med Assoc J*. 2014;186(14):1057–64.
15. Lau J, Ioannidis JP, Schmid CH. Quantitative synthesis in systematic reviews. *Ann Intern Med*. 1997;127(9):820.
16. Liberati A, Altman DG, Tetzlaff J, Murrow C, Gøtzsche PC, Clarke M, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *Ann Intern Med*. 2009;151(4):W65–94.
17. Matthew EF, Eleni EP, George AM, Georgios P. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *Fed Am Soc Exp Biol*. 2015;20 Sep 2007.
18. Mchugh ML. Interrater reliability: the kappa statistic. *Biochem Med*. 2012;22(3):276–82.
19. Moher D, Jadad AR, Nichol G, Penman M, Tugwell P, Walsh S. Assessing the quality of randomized controlled trials: an annotated bibliography of scales and checklists. *Control Clin Trials*. 1995;16:62–73.
20. Pae C-U. Why systematic review rather than narrative review? *Psychiatry Investig*. 2015;12(3):417.
21. Russell RM. Issues and challenges in conducting systematic reviews to support development of nutrient reference values: workshop summary. Rockville: U.S. Dept. of Health and Human Services, Agency for Healthcare Research and Quality; 2009.
22. Santos JRA. Cronbach's alpha: a tool for assessing the reliability of scales. *J Ext*. 1999;37:2.
23. Slim K, Nini E, Forestier D, Kwiatkowski F, Panis Y, Chipponi J. Methodological index for non-randomized studies (MINORS): development and validation of a new instrument. *ANZ J Surg*. 2003;73(9):712–6.
24. Torgerson C. *Systematic reviews*. London: Continuum; 2003.
25. Uman LS. Systematic reviews and meta-analyses. *J Can Acad Child Adolesc Psychiatry*. 2011;20(1):57–9.
26. Verhagen AP, Vet HCD, Bie RAD, Boers M, Brandt PAVD. The art of quality assessment of RCTs included in systematic reviews. *J Clin Epidemiol*. 2001;54(7):651–4.
27. Weil RJ. The future of surgical research. *PLoS Med*. 2004;1(1):e13.
28. Wright RW, Brand RA, Dunn W, Spindler KP. How to write a systematic review. *Clin Orthop Relat Res*. 2007;455:23–9.
29. Zeng X, Zhang Y, Kwong JS, Zhang C, Li S, Sun F, et al. The methodological quality assessment tools for preclinical and clinical studies, systematic review and meta-analysis, and clinical practice guideline: a systematic review. *J Evid Based Med*. 2015;8(1):2–10.



Kelsey L. Wise, Brandon J. Kelly,
Michael L. Knudsen, and Jeffrey A. Macalena

38.1 Reliability Studies

38.1.1 Introduction

Measurements and their reliability comprise an essential part of orthopedic research [5, 10, 13, 23, 24, 32, 36, 49]. The value of a measurement lies in its ability to be compared. There are no perfect measurement instruments, and each has a certain amount of error. Measurement error refers to how well a particular instrument performs within a given population. Less measurement error yields more precise data. The primary factor in determining acceptable measurement error is the expected range of measurements [22, 26].

Reliability refers to the reproducibility of a scale and is defined as the relationship between the expected distribution of measurements, the actual distribution of measurements, and the resulting measurement error [26, 34]. Reliability and agreement differ. If a test is performed and the result does not change regardless of rater, subject, or other variables, there would be perfect agreement [22]. However, this test is of little value to the clinician. Reliability evaluates the spectrum of measurements [22]. The difference between agreement and reliability is important because

reliability is better able to evaluate the usefulness of a scale, measure, or tool [22]. No survey or measurement instrument has perfect reliability. The reliability of measurements must be scrutinized when critically evaluating studies [26].

Reliability is vital for an instrument to be clinically useful and applicable. Once reliability has established validity, feasibility and acceptability must be assessed. Validity is characterized by how well the instrument achieves the intended measure. Feasibility refers to time, availability of required resources, sample size, and lack of fatal flaws in study design [34]. Acceptability is the utility of the instrument in clinical practice. A reliability study is only useful if the subjects, raters, and testing conditions in the study mirror, or are at least similar to, those variables in clinical practice or research [22].

Fact Box 38.1

Key definitions	
Term	Definition
Reliability	Reproducibility of scale
Agreement	Reproducibility of measurement
Validity	Instrument's ability to achieve intended measure
Feasibility	Resources, availability, time, sample size, study design flaws
Acceptability	Utility of instrument in clinical practice
Measurement error	Instrument's performance in given population

K. L. Wise · B. J. Kelly · M. L. Knudsen ·
J. A. Macalena (✉)
Department of Orthopaedic Surgery,
University of Minnesota, Minneapolis, MN, USA
e-mail: maca0049@umn.edu

38.1.2 Types of Reliability Studies

In a reliability study, investigators must choose the type of measurement, the tool used to obtain measurement, and how the measure will be used in a clinical or research setting [22]. Internal consistency reliability, test-retest reliability, intraobserver reliability, interobserver reliability, and alternate-form reliability are the most utilized measures in reliability studies [22].

38.1.2.1 Internal Consistency

Internal consistency indicates the uniformity of a scale, measuring the similarity of an individual's responses across several items [34]. It is a gauge of how well different items measure the same thing [26]. Cronbach's coefficient alpha is a numeric value of internal consistency [26]. It is a statistic that illustrates the homogeneity of the scale [26]. It is derived by associating the score range of each component scale with each independent observer score and then comparing to the total variance of all items.

Clinical Vignette 1

Dr. Sessions measures the internal consistency of a newly formulated survey that includes five questions about the quality of life of patients with below-knee amputations. The hypothesis is that individuals of similar age, gender, mental functioning, and pain ratings have comparable level of function, self-image, mental health, and pain ratings at a mean follow-up of 3 years.

For responses, yes = 1 and no = 0.

Patient	Item 1	Item 2	Item 3	Item 4	Item 5	Summed scale score
1	1	0	1	1	1	4
2	0	1	0	1	0	2
3	1	1	1	1	1	5
4	0	1	0	1	0	2
5	1	1	1	0	1	4
Percentage positive	$\frac{3}{5} = 0.6$	$\frac{4}{5} = 0.8$	$\frac{3}{5} = 0.6$	$\frac{4}{5} = 0.8$	$\frac{3}{5} = 0.6$	

The sample mean and sample variances are calculated first.

The sample mean is:

$$(4 + 2 + 5 + 2 + 4) / 5 = 3.4$$

The sample variance is:

$$(4 - 3.4)^2 + (2 - 3.4)^2 + (5 - 3.4)^2 + (2 - 3.4)^2 + (4 - 3.4)^2 / 5 = 1.44$$

Coefficient alpha for a series of dichotomous items is

$$\left[\frac{k}{k-1} \right] \left[1 - \frac{\sum_1^k (\%Positive)(\%Negative)}{\sigma^2} \right]$$

where k is the number of items in the scale.

Coefficient alpha is

$$\left[\frac{5}{5-1} \right] \left[1 - \frac{(0.6)(0.4) + (0.8)(0.2) + (0.6)(0.4) + (0.8)(0.2) + (0.6)(0.4)}{1.44} \right] = 0.35$$

The internal consistency, measured with Cronbach's coefficient alpha, is 0.35, indicating a very poor correlation between items.

For group comparison, reliability statistics measured should exceed 0.70, and for individual comparison, they should exceed 0.90 [34]. Internal consistency can be improved by adding questions or measurements, clarifying existing questions to reduce response bias, and reducing variability in the respondents [26].

A major drawback of internal consistency is that it only requires a single administration of an instrument; thus possible differences between raters, timing, and situation can result in measurement error [22]. Test-retest reliability, intraobserver reliability, and interobserver reliability mitigate this source of measurement error by incorporating multiple measurements into the calculations [22].

Fact Box 38.2

Internal consistency reliability statistic interpretation

Population	Cronbach's coefficient alpha	Interpretation
Individual comparison	≥0.90	Acceptable agreement
Group comparison	≥0.70	Acceptable agreement

38.1.2.2 Test-Retest

Test-retest reliability measures response stability over time [34]. Participants respond to a survey at two distinct time points to measure response stability. Subsequently, correlation coefficients (*r* values) are determined to compare responses [26, 34]. *R* values are considered strong if they equal or exceed 0.70 [26]. A source of error in this study is a change in characteristics of the test setting with multiple administrations.

Clinical Vignette 2

Dr. Siljander wishes to evaluate postoperative pain scores using a visual analog scale (VAS) in patients undergoing open reduction internal fixation (ORIF) of the distal radius. Fifteen patient responses are assessed at 2 and 4 h postoperatively. Responses to the VAS are scaled 0–100, and the response at 2 h postoperatively (time 1) is compared to 4 h postoperatively (time 2). The correlation coefficients of the two sets of data are compared. The *r* value is calculated to be 0.98, which indicates an excellent test-retest reliability at 2 and 4 h postoperatively.

	Time 1	Time 2
Patient 1	71	68
Patient 2	75	70
Patient 3	78	70
Patient 4	84	83
Patient 5	81	77
Patient 6	85	88
Patient 7	90	83
Patient 8	22	21
Patient 9	44	36
Patient 10	52	48
Patient 11	50	44
Patient 12	66	61
Patient 13	38	30
Patient 14	84	82
Patient 15	85	73
Σ	1005	934

Correlation coefficient, *r*:

$$r = \frac{n[\sum(\text{Time 1})(\text{Time 2})] - [\sum(\text{Time 1})][\sum(\text{Time 2})]}{\sqrt{[n\sum(\text{Time 1}^2) - \sum(\text{Time 1})^2][n\sum(\text{Time 2}^2) - \sum(\text{Time 2})^2]}}$$

r = 0.98

38.1.2.3 Intraobserver

Intraobserver reliability, while similar to test-retest reliability, measures reliability of measurements on the same data. This test has been used in

the evaluation of imaging modalities [17, 35, 46]. Intraobserver reliability tends to result in a higher reliability compared to the test-retest and interobserver study designs, as time is the only variable.

Clinical Vignette 3

Dr. Todhunter desires to evaluate the intraobserver reliability of three-dimensional computed tomography (3D CT) for measuring femoral torsion. He performs this measurement five times a week, for 5 weeks, on the same imaging study. Dr. Todhunter calculates the variance between all 25 images to be 3.72° and variance between weekly samples to be 5.30°. With an instrument-recorded measurement error of 1.33°, he derives an intraobserver reproducibility of 0.96:

$$\frac{\text{between_subjectSD}^2 + \text{between_observerSD}^2}{\text{between_subjectSD}^2 + \text{between_observerSD}^2 + \text{measurement_errorSD}^2}$$

$$\frac{(3.72)^2 + (5.30)^2}{(3.72)^2 + (5.30)^2 + (1.33)^2} = 0.96$$

38.1.2.4 Interobserver

Interobserver reliability measures agreement of two or more raters performing the same measurements. This is the best test for reliability, as this is the broadest of the five. When interobserver reli-

ability is high, this test is able to stand alone. However, with low interobserver reliability, it can be helpful to include test-retest reliability and/or intraobserver reliability to provide information on potential sources of error.

Clinical Vignette 4 (Continued from Previous Vignette)

Dr. Todhunter also evaluates the interobserver reliability of 3D CT femoral torsion measurements by having ten of his colleagues perform this measurement five times. He then compares each of their set of five measurements with his final weeks' set of five measurements. He calculates a variance of 3.11° among all 55 images, 1.11° between the 11 observers, and 1.33° for the instrument-recorded measurement error. He denotes an interobserver reliability measure of 0.76.

$$\frac{\text{between_subjectSD}^2}{\text{between_subjectSD}^2 + \text{between_observerSD}^2 + \text{measurement_errorSD}^2}$$

$$\frac{(3.11)^2}{(3.11)^2 + (1.11)^2 + (1.33)^2} = 0.76$$

38.1.2.5 Alternate Form

Alternate-form reliability offers a solution to the problem of the practice effect, as it involves rewording items to measure the same element. Items are administered to the

same group of individuals at different times, and correlation coefficients are calculated. Changing the order of answer choices in a survey is a simple way to test alternate-form reliability [26].

Clinical Vignette 5

Dr. Williamson investigates how often patients performed their home physical therapy exercises after total shoulder arthroplasty. Two different surveys are utilized.

Item 1: How many times per day did you complete home exercises in the last week?

1. 1 time per day
2. 2 times per day
3. 3–4 times per day
4. 5–8 times per day.

Item 2: During the last week, how often did you complete daily home exercises?

1. Every 24 h.
2. Every 12 h.
3. Every 6–8 h.
4. Every 2–5 h.

Fact Box 38.3

Reliability studies		
Types of reliability	Measure	Statistic
Internal consistency	Homogeneity of scale	Cronbach's coefficient alpha
Test-retest	Stability of scale	Correlation coefficient (<i>r</i> value)
Intraobserver	Response stability of individual rater	Intraclass correlation coefficient
Interobserver	Response stability of separate raters	Intraclass correlation coefficient
Alternate form	Stability of response to survey variation	Intraclass correlation coefficient

38.1.3 Participant Selection and Rating Assignment**38.1.3.1 Rater Selection**

Individual raters can introduce variation in measurement. In a self-administered survey, the rater is also the subject. When more than one rater contributes measurements, there are two points of variability, rater expertise and rater practice setting. In general, more reliable ratings are seen with more experienced raters. However, diversity among raters with regard to level of training and practice setting is valuable in situations where raters using the measuring tool have different levels of expertise and practice settings. Rater expertise and practice setting must be disclosed in reliability studies [22].

38.1.3.2 Subject Selection

Study subjects should be representative of clinical practice. To strengthen reliability, include subjects with a wide range of clinical conditions or variables for the measurement focus [22]. Homogenous groups will have stronger raw agreement, whereas heterogeneous groups will have stronger reliability because of increased sample variability [22]. Reliability is increased with greater subject variability relative to the measurement error [22].

Fact Box 38.4**Participant selection****Rater selection**

- Rater expertise and practice setting should reflect audience
- Rater expertise improves reliability

Subject selection

- Heterogeneous groups have stronger reliability
- Homogenous groups have stronger raw agreement

38.1.4 Data Evaluation

There are many techniques and statistical tests to describe or measure reliability. Calculating or reporting more than one statistical estimate may be beneficial.

38.1.4.1 Categorical Data

Categorical data are discrete and qualitative variables. For example, categories are nominal (e.g., “absent,” “present”) or ordinal (e.g., “mild,” “moderate,” or “severe”) [43].

38.1.4.2 Continuous Data

In some situations, ratings are on a continuous scale. Examples would be hip range of motion or the number of steps a patient can take.

38.1.4.3 Kappa Coefficient

The *kappa coefficient* is used for categorical data. It measures agreement and compares observed agreement with possible agreement beyond chance [12, 43]. Kappa is the most frequently reported statistic in reliability studies examining orthopedic fractures [4]. It takes the form:

$$k = \frac{\text{observed agreement} - \text{chance agreement}}{1 - \text{chance agreement}}$$

In symbols, this is:

$$k = \frac{P_O - P_C}{1 - P_C}$$

P_O is the proportion of observed agreements, and P_C is the proportion of agreements expected by chance [43].

The value for perfect agreement is 1.0, and the value for no agreement beyond chance is 0.0. Negative values indicate that the measured agreement is worse than chance alone [22]. Kappa is often used in interobserver reliability (two or more clinicians rate same patient) or intraobserver reliability (single clinician rates same patient two or more times). Kappa is not as useful when skewed distribution of responses exists because agreement above chance is less likely.

38.1.4.4 Phi Statistic

The phi statistic can also be used for categorical data, and it has the benefit of measuring agreement

independent of chance [22]. Despite being useful when the distribution of responses is skewed, it is used infrequently.

38.1.4.5 Pearson Correlation

The *Pearson correlation coefficient* (PCC), also referred to as Pearson’s “ r ,” measures linear correlation with a range of +1 (perfect correlation) to –1 (perfect but negative correlation) and 0 value signifying no relationship [2]. It is often used with continuous data. The limitation of Pearson correlation is that two groups of measurements may still be in poor agreement, despite perfect correlation. As a result, in reliability studies, the Pearson correlation may poorly describe two-variable relationships.

38.1.4.6 Intraclass Correlation Coefficient

Reliability represents the reproducibility of a scale, and *intraclass correlation coefficients* (ICC) determine values closest to this definition of reliability. Intraclass correlation coefficients are calculated through repeated measures of analysis of variance [16]. The intraclass correlation coefficient is statistically defined as the ratio of the variance of interest over the sum of variance plus error [41]. Continuous and categorical data can be analyzed with the intraclass correlation coefficient, but it is most commonly used with continuous data.

Fact Box 38.5

Data evaluation		
Data types	Variables	Example
Categorical data	Discrete or qualitative	“Absent” or “present”
Continuous data	Numeric or quantitative	Radiographic measurements
Statistic	Ideal data type	Measurement
Kappa coefficient	Categorical data	Agreement
Phi statistic	Categorical data	Agreement independent of chance
Pearson correlation	Continuous data	Linear correlation
Intraclass correlation coefficient (ICC)	Continuous (most common) or categorical data	Reproducibility

38.1.5 Sample Size

The number of raters and subjects is controlled in reliability studies. To increase precision, increasing the number of subjects carries more weight than increasing the number of raters (particularly when there are more than four raters) [22]. However, increasing either subjects or raters will narrow the confidence interval. In an effective study design, the number of raters is selected first based on study generalizability and feasibility, and then subject number for desired precision is selected [22].

Generalizability of the raters is determined by their characteristics. Feasibility of the raters or ratings depends more on the subjects. Radiographs can be rated multiple times, whereas patients may not tolerate more than one examination. Sample size calculations are performed once the number of raters is decided.

For intraclass correlation coefficients, the number of subjects is determined by the minimum acceptable reliability [48]. Alternatively, the desired precision of the reliability estimate can be used to determine subject number [18].

38.1.6 Interpretation of Results

The majority of reliability statistics use a 0.0–1.0 scale. A value of 1.0 indicates all variability is due to true subject differences; a value of 0.0 indicates all variability is due to error [22]. Table 38.1 shows the agreement associated with certain kappa values. In reality, most studies have reliability that falls between 0.3 and 0.7 [4]. Ultimately, readers must determine how applicable study design, raters, subjects, and measurement tools are to their practices.

Table 38.1 Interpretation of the agreement with kappa coefficient

Kappa value	Agreement
0	Absent
>0–0.20	Slight
0.21–0.40	Fair
0.41–0.60	Moderate
0.61–0.80	Good
0.81–1.00	Excellent

38.1.7 Summary

Reliability testing of scales and items is a valuable tool in orthopedic surgery, as it provides quantitative data on the performance of an instrument [26]. Types of reliability studies include internal consistency, test-retest, intraobserver, interobserver, and alternate form. Categorical data typically use the kappa statistic, while continuous data tend to utilize the intraclass correlation coefficient. A correlation coefficient, or r value, with a value of 0.70 or greater is generally accepted and indicates good reliability.

38.2 Surveys

38.2.1 Introduction

Surveys are useful instruments in orthopedic research for gathering data on the demographics, practices, or ideas of a group of individuals at a specific time or to compare changes over time [11, 14, 21, 33, 38, 44]. Constructing surveys involves writing questions that may ultimately be converted into numbers to allow for statistical analysis [30]. The results of surveys give insight into the adoption of existing literature and help guide clinical practice and future research [44]. Surveys are administered by telephone, mail, fax, or electronically.

Surveys allow for convenient and inexpensive research that can cover a large population over a short period of time [21]. Surveys may be the only method of conducting certain types of research, such as understanding current viewpoints or attitudes of orthopedic surgeons. Additionally, surveys may act as preliminary studies for successive research ideas [21].

Survey validity is jeopardized by low response rates. Response rates have been as low as 15% among surgeons [44]. This has been attributed to various factors, including busy work schedules, an increasing number of commercial requests, increased paperwork, and low priority [6, 7, 25, 40, 44]. Low responses can result in nonresponse bias, as individuals who respond can have meaningful differences from those who do not. This

can lead to a nonrepresentative sample and a subsequent loss of survey credibility [44]. Biases tend to be limited by achieving response rates of at least 70% [33]. In addition to increasing the validity of surveys, a higher response rate may also eliminate the costs needed for resending surveys [37].

A systematic approach to survey design, development, testing, administration, and considering strategies for elevating response rate are imperative to decrease bias and increase generalizability [11].

38.2.2 Survey Design

38.2.2.1 Determining the Objective

A well-defined objective is important for a successful survey [11]. This involves careful consideration of the topic and target audience [21]. Strong research questions are specific, simple, meaningful, interesting, and answerable [33].

38.2.2.2 Develop the Survey Instrument

After a clearly stated research question has been formulated, investigators may modify an existing survey or develop a new instrument [33]. The survey instrument acts as the interface between study objectives and responses of individuals surveyed [21]. Strategies to identify item content include conducting focus groups of potential subjects, discussing with an expert panel, or utilizing the Delphi technique, a process by which items are selected and ranked by a group of experts until a consensus is reached [33]. Regardless of which strategy is employed, a team of colleagues should be involved in the development process to lend face validity to the instrument before it is tested.

38.2.2.3 Identifying the Sampling Frame

It is difficult to administer a survey to all individuals within a goal population as a result of the population size or the hardship of identifying and contacting all possible respondents [11, 38]. As a result, a fraction of the target population is sur-

veyed. The “sampling frame” comprises the target population for the survey, while the “sampling element” includes the respondents from whom data is gathered and evaluated [1].

Sample selection may be random (probability design) or nonrandom (nonprobability design) [1]. Probability designs include simple random sampling, systematic random sampling, stratified sampling, and cluster sampling [11].

- *Simple random sampling*: Each person has the same probability of being selected. Selection takes place through techniques such as a lottery process or a random-number generator [45]. This technique requires little prior knowledge of a population but may not capture certain groups or be efficient [1].
- *Systematic random sampling*: A starting point is arbitrarily chosen on a list. Subjects are then selected in a methodical manner (e.g., every fifth subject) [11]. This technique has high precision and allows ease of analyzing data and measuring sampling errors [1]. However, the ordering of the list of subjects in the sample frame can create biases, certain groups may be excluded, and there may be a lack of efficiency [1].
- *Stratified random sampling*: Possible respondents are divided into defined groups. Within these groups, individuals are randomly sampled by either simple or systematic sampling [11]. This technique allows disproportionate sampling to be possible and has high precision [1]. A disadvantage is that the advanced knowledge of a population can result in more complex analysis.
- *Cluster sampling*: The population is broken down into heterogeneous clusters, and specific clusters are sampled [11]. This method has lower costs and allows sampling of groups if individuals are not available [1]. However, this method is more complex when analyzing data and evaluating sample errors, and it has low precision.

In nonprobability sampling, individuals in a population have an unequal chance of being involved in the sample [11]. Nonprobability sam-

pling designs include purposive sampling, quota sampling, chunk sampling, and snowball sampling [1].

- *Purposive sampling*: Selected participants meet specific criteria (e.g., they are hand surgeons) [11].
- *Quota sampling*: A selected number of respondents are selected based on specific characteristics (e.g., a researcher sets a quota of 20% female spine surgeons between the ages of 30 and 50 for the sample) [11].
- *Chunk sampling*: Participants are chosen based on convenience (e.g., attendees of a conference) [11].
- *Snowball sampling*: Investigators find individuals who meet certain criteria, who then go and find more individuals who meet the criteria [11].

The degree of generalizability of a survey is determined by similarity of respondents to nonrespondents [11]. It is rarely feasible to obtain data on how respondents and nonrespondents differ. The best way to overcome this setback is to strive for a high response rate [11].

Sample selection		
Sample selection type	Design (probability or nonprobability)	Features
Cluster sampling	Probability	Population separated into heterogeneous clusters and specific clusters are sampled
Purposive sampling	Nonprobability	Criteria-based selection
Quota sampling	Nonprobability	Select number of respondents based on specific qualities
Chunk sampling	Nonprobability	Individual availability-based selection
Snowball sampling	Nonprobability	Selected individuals who meet the criteria select more participants

38.2.3 Development

38.2.3.1 Question Stems

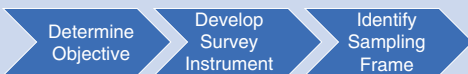
Each question stem included in a survey should be relevant to the overarching purpose of the survey and ideally relate to recent events or common knowledge to produce high-quality survey data [44]. Questions must be succinct and clear and contain 20 or fewer words [11, 33]. Question types that should be avoided include double-barreled questions (combining two questions), the halo effect (referencing influential sources that can influence a response), loaded questions, double negatives, and a question with no comparator, modifiers (i.e., “almost everyone” or “usually”), or complex vocabulary [33, 38, 42]. Other techniques for providing proper question stem structure include using complete sentences and softening the impact of potentially controversial questions [1, 44].

38.2.3.2 Question Responses

Responses should be concise and impartial. Response formats can be either open or closed [11, 33]. Open responses allow subjects to answer questions in free text. This format offers the advantage of minimizing limitations in responses. However, free responses are more time-consum-

Fact Box 38.6

Study Design



Sample selection		
Sample selection type	Design (probability or nonprobability)	Features
Simple random sampling	Probability	All individuals equal chance of selection
Systematic random sampling	Probability	Individuals selected at predetermined interval from list
Stratified random sampling	Probability	Individuals categorized and randomly sampled through simple or systematic random sampling

ing, and results are more difficult to convert into analyzable data [33]. Closed response formats include nominal, ordinal, interval, binary, and ratio measurements [11].

- *Nominal responses*: These responses involve a list of mutually exclusive items (e.g., physicians, nurses, medical students) that reflect qualitative distinctions [11].
- *Ordinal responses*: These responses imply a ranked order. The traditional Likert scale, with answers ranging from “strongly disagree” to “strongly agree,” is commonly used in surveys [33]. This format allows the grouping of like-minded attitudes and viewpoints. An odd number of points allows a neutral response, while an even number of points forces a commitment [33]. In a phenomenon called the “floor” or “ceiling” effect, respondents tend to choose responses clustered at the bottom or top of a scale. Increasing the number of points on a scale may exacerbate this phenomenon [33].
- *Interval and ratio measurements*: These measurements show continuous responses. A ratio scale has an absolute zero (e.g., height and weight), while intervals do not have a true zero (e.g., interval time of day; 1 PM to 2 PM is the same as 5 PM to 6 PM because both are 1 h increments).

It is helpful to involve a biostatistician to ensure that answer choices are in a format where the data can easily be analyzed.

Question responses that should be avoided include vague answer choices (e.g., “other” or “unknown”), absolute terms (e.g., “always” or “never”), abbreviations, complex answers, and asking respondents to rank responses [11, 21]. There should be an equal number of positive and negative option choices for scaled questions.

38.2.3.3 Chronology

The chronology of the questions can have a major influence on the response rate of surveys. It is helpful to begin with demographic questions, as these are simple and nonthreatening [33]. The first non-demographic question should be clearly

stated, the most applicable, and noteworthy [44]. Items can be grouped based on content to help with respondents’ thought process and memory [33]. The filter technique allows responders to omit questions based on prior responses, thus limiting frustration with irrelevant or non-applicable questions [50].

38.2.3.4 Survey Length

Surveys should be succinct. One study found significantly lower response rates to a threshold level of 1000 words [20]. Though this threshold was not directly aimed at orthopedic surgeons, this study suggests that response rate may be considerably affected by minor modifications in questionnaire length.

38.2.3.5 Survey Format

The format of a survey is a vital part of presentation. There should be explicit instructions throughout, as opposed to solely at the beginning of surveys. Arrows or symbols to provide visual navigation are beneficial for assistance with survey completion [44]. The size and style of the font should be easy to read [11]. Emphasizing important words or phrases, numbering questions, providing appropriate spacing, and listing answers vertically instead of horizontally are all helpful techniques [1, 44]. It is helpful to create a polished, yet unique, appearance that allows a survey to stand out from other questionnaires responders may receive.

38.2.3.6 Cover Letters

Cover letters provide the initial opportunity to persuade readers to complete surveys [44, 50]. The letter should clearly state the purpose of the survey and why participants were selected [33]. Confidentiality and optional completion of the survey should be conveyed [1, 50]. Details should be included regarding when the survey should be completed and who to contact if participants have questions. Finally, the cover letter should contain an affirmation that the recipient’s participation is vital to the success of the survey and an expression of gratitude to the recipient for his or her time [37].

Fact Box 38.7

Survey development	
Survey parts	Features
Question stem	Succinct, clear, avoid objectionable questions
Question response	“Open” (free text) or “closed” (structured)
Chronology	Demographics first, personal questions last
Survey length	Concise (as short as possible to convey idea)
Survey format	Polished and unique with instructions throughout
Cover letter	Recruits and attracts participants

38.2.4 Pilot Testing

Pilot testing allows trial on methods of sampling, data collection, and survey administration. Investigators ask colleagues or other participants who are similar to desired respondents to evaluate questions and provide feedback [11]. This is important to evaluate for mistakes, gauge the time needed for completion, understand whether the survey conveys the proposed message, and test to see if the survey grabs the reader’s interest [50].

38.2.5 Methods of Survey Administration

Surveys can be given over the phone, through postal mail, via fax, or electronically. The selected administration technique depends on the targeted audience, the type of information desired, financial limitations, and whether test properties were constructed.

38.2.5.1 Telephone

Telephone surveys offer the advantage of increasing accuracy, as the administrator can ensure question comprehension and complete response from the responder. Downfalls include the expense, difficulty in achieving a high response rate, and the susceptibility to distortion and interview bias [33, 44].

38.2.5.2 Postal

Postal delivery is the traditional method of survey administration. Postal surveys allow responders to complete questionnaires in privacy, and interviewer distortion is minimized [6, 44, 50]. Postal surveys may also increase validity by giving responders time to formulate answers and utilize various sources [44]. Disadvantages of postal surveys include the cost of supplies, the time necessary for the labor of collecting responses, and the potential decrease in accuracy when manually recording responses as opposed to responses automatically being entered into a database [29, 44].

38.2.5.3 Fax

Respondents of fax surveys may return the survey through fax or by postal mail. Data can be collected manually or with utilization of optical character recognition by some fax machines [29]. Character recognition may enhance the accuracy of the recorded data. Costs of fax surveys are similar to costs of postal surveys [44].

38.2.5.4 Electronic

The current mainstay of survey administration is through electronic distribution [9, 21, 25, 47]. Electronic surveys may be web-based, where respondents fill out a questionnaire on a website. Alternatively, surveys can be conducted through email, as an attachment, or embedded in the text of the email.

Advantages of electronic surveys include ease of completion, ability to reach a large audience, cost-effectiveness, and the immediate availability of responses [9, 21, 25, 29, 44, 47]. Simple descriptive statistics usually are embedded, allowing concurrent analysis for researchers, while more complex statistical analysis is generally accomplished by exporting data to statistical software [47]. This reduces time and resources, as well as the possibility of human error. Electronic surveys allow control of question order, thus preventing respondents from changing prior answers [9]. Finally, email surveys have been showed to be associated with a lower number of unanswered questions [29].

There are many possible sources of weakness with electronic surveys. Potential selection bias is a challenge, as online surveys may not necessarily be suitable for certain groups of participants. This could lead to an inaccurate or incomplete cross section of the population [30]. Email surveys may be dismissed as junk mail [29, 47]. It can be difficult accessing desired mailing addresses. McPeake et al. [30] found that almost 10% of emails sent using a 1-year-old contact list were returned as undeliverable.

VanDenKerkhof and colleagues [47] discuss the following strategies as methods to aid with ease of online surveys:

1. Provide drop-down lists with “select one” as the default option rather than an actual response, as the latter may create bias.
2. Even if only one response is required, participants should be given the option of selecting multiple responses or providing a comment in a text box.
3. A progression bar should be provided for respondents to gauge progress.
4. One question per screen simplifies the visual field and makes the survey user-friendlier.
5. With a rising number of email surveys, response can increase by sending an initial contact (i.e., a postcard).

Overall, the accessibility and quick data collection of electronic surveys make these favorable options for survey administrators. However, postal and fax surveys have been reported to have higher response rates than telephone or electronic surveys [19, 25, 28, 29]. Possible reasons for lower response rates with electronic surveys include lack of familiarity with the Internet, inconsistent Internet access, lack of trust submitting confidential information over the Internet, and survey saturation [30]. Using mix-mode designs can lead to increased response rate [1]. For example, an electronic survey could be distributed first, and a postal mail survey could be utilized for the second distribution.

Fact Box 38.8

Methods of survey administration		
Method	Advantage	Disadvantage
Telephone	Increased accuracy of response	Difficult to obtain high response rate
Postal	Confidentiality, anonymity, convenience, minimize interviewer distortion	Time to collect, cost of supplies, decreased accuracy from manual entry
Fax	Confidentiality, anonymity, convenience, minimize interviewer distortion	Time to collect, decreased accuracy from manual entry
Electronic	Ease, cost, immediate responses	Selection bias, junk mail

38.2.6 Response Rate

High response rates increase validity, enhance the precision of parameter estimates, and reduce the risk of selection bias. Low response rates increase the chance that respondents and nonrespondents differ in a meaningful way, which may undermine the results of survey responses [11]. Investigators can report the *actual response rate* or the *analyzable response rate*. The *actual response rate* includes respondents with partially completed questionnaires and opt-out responses, mirroring the sampling element. However, the *analyzable response rate* reflects a proportion of the sampling frame based on partial or full completion of questionnaires [11]. A response rate of at least 70% has been viewed as acceptable for external validity [50]. However, response rates between 60% and 70% and sometimes less than 60% (e.g., for controversial topics) may be acceptable [42]. There are a number of techniques that enhance the response rate of surveys.

38.2.6.1 Convenience

Orthopedic surgeons are busy individuals; thus enhancing the convenience of survey participation is crucial for increasing participation. The estimated time for completion of a survey should be included in the cover letter (i.e., “less than 10

min”) [30]. For email surveys, embedding the link of the survey in the text rather than including the survey as an attachment is likely to increase response rate. If utilizing postal surveys, providing stamped return envelopes, as opposed to franked return envelopes or no envelope at all, is a simple, low-cost, and effective method of increasing response rate [40].

38.2.6.2 Precontacts

Informing participants of an upcoming survey may enhance questionnaire participation. There have been mixed reports on increasing response rate with prenotification letters [1, 40, 50], though it is thought that a prenotification is particularly important with email surveys for these questionnaires to avoid being viewed as “junk mail” and deleted [47]. A prenotification letter should be personalized, concise, and positive. It should also aim at building anticipation as opposed to providing excessive detail about the survey [1].

38.2.6.3 Multiple Contacts

Dillman et al. described a method of contacting nonrespondents numerous times with the intention of increasing response rate [1]. Each approach is slightly different in this method, allowing multiple opportunities to catch more responders. For example, multiple attempts may be made with a mail survey before attempting a telephone survey.

38.2.6.4 Incentives

A multitude of incentives, such as money or prizes, have been used with the aim of increasing survey completion [3, 8, 19]. The efficacy of incentives does not have strong support in the literature; thus the cost-effectiveness of this strategy is questionable [44].

38.2.6.5 Reminders

Reminders can increase response rates to surveys [31]. With postal surveys, each additional mailed reminder may increase response up to 30–50% from the initial response rate [42]. Dillman and colleagues [1] proposed sending up to 3 follow-up reminders for nonrespondents—1 sent a week after the initial mailing and 2 sent between 3 and 7 weeks after the initial mailing. The use of reminders has also been shown to be effective with

electronic surveys [15, 39]. If known, the current response rate should be included in each reminder, acting as an incentive for readers to participate.

38.2.6.6 Principal Investigator Contact

Finally, a personal gesture that may increase response is having the principal investigator (PI) reach out to subjects directly.

Fact Box 38.9

How to enhance response rate?

- Convenience: decrease time and increase availability
- Precontacts: build anticipation and awareness
- Multiple contacts: Dillman approach [35] for Internet and postal surveys
- Incentives: infrequently cost-effective but may increase responses
- Reminders: increased responses in dose-dependent relationship
- PI contact subject her/himself

38.2.7 Reporting Surveys

Study findings are reported with an introduction, methods, results, and discussion. The hypothesis should be clearly stated. A brief summary of pretesting, instrument revision, and method administration should be included. It is very important to include the response rate. If applicable, commentary should be given on how respondents and nonrespondents differ, limitations, and sources of potential bias [21]. Information gathered from the survey and how this data relates to the original hypothesis should be discussed. Finally, significance of results should be highlighted, as well as directions for future research and implications for research, education, and clinical care.

38.2.8 Costs Associated with Surveys

Survey fees include labor costs, pretesting, survey administration, repeat survey administration for nonresponders, and data analysis [44]. Some of these costs are eliminated with the utilization of Internet surveys.

38.2.9 Ethical Considerations

Research should be approved by an ethics review committee prior to the administration of a survey [27]. The voluntary nature of a survey and confidentiality should be emphasized in the cover letter [42]. Confidentiality can be obtained by using codes, destroying questionnaires quickly after data collection, or removing identifying information on the questionnaire. Furthermore, participants should be given comprehensive information regarding the purpose of the study, the research sponsors, and who to contact for further questions [1, 27].

38.2.10 Limitations of Surveys

Surveys are limited as research tools. They rely on respondents' memory, honesty, and understanding of the survey text, all components which are difficult to measure [21]. Surveys, by nature, can be restrictive, unless interviews or open-ended questions are used. However, free responses and personal interview have the setback of more difficulty with the conversion of data into an analyzable form. While surveys can establish a relationship between variables, a further downside is the inability of surveys to establish causality [21]. Finally, surveys are limited by the response rate, which can potentially cast doubt on the accuracy of survey data.

38.2.11 Summary

A survey is a relatively easy, quick, and inexpensive form of research that has the capability of reaching a large population of individuals. Surveys directed at orthopedic surgeons can provide valuable contributions to the literature, as they assist in furthering understanding of current orthopedic ideas and practices. Each stage of survey development and administration should be carefully planned to ensure study objectives are answered, bias is minimized, and generalizability is increased.

Take-Home Message

- Reliability studies gauge the reproducibility of a measuring instrument, while surveys allow rapid and cost-effective data collection of a large population. An understanding of the components and administration of both studies will allow optimal utilization and appraisal of these study types within orthopedic literature.

References

1. Aday LA, Cornelius LJ. Designing and conducting health surveys: a comprehensive guide. 3rd ed. San Francisco: Jossey-Bass; 2006.
2. Adler J, Parmryd I. Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry A*. 2010;77(8):733–42.
3. Asch DA, Christakis NA, Ubel PA. Conducting physician mail surveys on a limited budget. A randomized trial comparing \$2 bill versus \$5 bill incentives. *Med Care*. 1998;36(1):95–9.
4. Audigé L, Bhandari M, Kellam J. How reliable are reliability studies of fracture classifications? A systematic review of their methodologies. *Acta Orthop Scand*. 2004;75(2):184–94.
5. Avery DM, Matullo KS. Distal radial traction radiographs: interobserver and intraobserver reliability compared with computed tomography. *J Bone Joint Surg Am*. 2014;96(7):582–8.
6. Baron G, De Wals P, Milord F. Cost-effectiveness of a lottery for increasing physicians' responses to a mail survey. *Eval Health Prof*. 2001;24(1):47–52.
7. Bergk V, Gasse C, Schnell R, Haefeli WE. Mail surveys: obsolescent model or valuable instrument in general practice research? *Swiss Med Wkly*. 2005;135(13–14):189–91.
8. Bhandari M, Devereaux PJ, Swiontkowski MF, Schemitsch EH, Shankardass K, Sprague S, et al. A randomized trial of opinion leader endorsement in a survey of orthopaedic surgeons: effect on primary response rates. *Int J Epidemiol*. 2003;32(4):634–6.
9. Braithwaite D, Emery J, De Lusignan S, Sutton S. Using the Internet to conduct surveys of health professionals: a valid alternative? *Fam Pract*. 2003;20(5):545–51.
10. Bruinsma WE, Guitton TG, Warner JJP, Ring D, Science of Variation Group. Interobserver reliability of classification and characterization of proximal humeral fractures: a comparison of two and three-dimensional CT. *J Bone Joint Surg Am*. 2013;95(17):1600–4.

11. Burns KEA, Duffett M, Kho ME, Meade MO, Adhikari NKJ, Sinuff T, et al. A guide for the design and conduct of self-administered surveys of clinicians. *CMAJ*. 2008;179(3):245–52.
12. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. *Psychol Bull*. 1968;70(4):213–20.
13. Corona J, Sanders JO, Luhmann SJ, Diab M, Vitale MG. Reliability of radiographic measures for infantile idiopathic scoliosis. *J Bone Joint Surg Am*. 2012;94(12):e86.
14. Duffett M, Burns KE, Adhikari NK, Arnold DM, Lauzier F, Kho ME, et al. Quality of reporting of surveys in critical care journals: a methodologic review. *Crit Care Med*. 2012;40(2):441–9.
15. Fischbacher C, Chappel D, Edwards R, Summerton N. Health surveys via the Internet: quick and dirty or rapid and robust? *J R Soc Med*. 2000;93(7):356–9.
16. Fisher R. *Statistical methods for research workers*. 5th ed. Edinburgh: Oliver and Boyd Ltd.; 1925.
17. Gaumétou E, Quijano S, Ilharberborde B, Presedo A, Thoreux P, Mazda K, et al. EOS analysis of lower extremity segmental torsion in children and young adults. *Orthop Traumatol Surg Res*. 2014;100(1):147–51.
18. Giraudeau B, Mary JY. Planning a reproducibility study: how many subjects and how many replicates per subject for an expected width of the 95 per cent confidence interval of the intraclass correlation coefficient. *Stat Med*. 2001;20(21):3205–14.
19. Hocking JS, Lim MSC, Read T, Hellard M. Postal surveys of physicians gave superior response rates over telephone interviews in a randomized trial. *J Clin Epidemiol*. 2006;59(5):521–4.
20. Jepson C, Asch DA, Hershey JC, Ubel PA. In a mailed physician survey, questionnaire length had a threshold effect on response rate. *J Clin Epidemiol*. 2005;58(1):103–5.
21. Jones D, Story D, Clavisi O, Jones R, Peyton P. An introductory guide to survey research in anaesthesia. *Anaesth Intensive Care*. 2006;34(2):245–53.
22. Karanicolas PJ, Bhandari M, Kreder H, Moroni A, Richardson M, Walter SD, et al. Evaluating agreement: conducting a reliability study. *J Bone Joint Surg Am*. 2009;91(Suppl 3):99–106.
23. Lee KM, Chung CY, Park MS, Lee SH, Cho JH, Choi IH. Reliability and validity of radiographic measurements in hindfoot varus and valgus. *J Bone Joint Surg Am*. 2010;92(13):2319–27.
24. Lee KM, Lee J, Chung CY, Ahn S, Sung KH, Kim TW, et al. Pitfalls and important issues in testing reliability using intraclass correlation coefficients in orthopaedic research. *Clin Orthop Surg*. 2012;4(2):149–55.
25. Leece P, Bhandari M, Sprague S, Swiontkowski MF, Schemitsch EH, Tornetta P, et al. Internet versus mailed questionnaires: a controlled comparison (2). *J Med Internet Res*. 2004;6(4):e39.
26. Litwin MS. How to measure survey reliability and validity. In: Litwin MS, editor. *How to measure survey reliability and validity*. Thousand Oaks: Sage; 1995. p. 5–32.
27. Mailey SK. Increasing your response rate for mail survey data collection. *SCI Nurs*. 2002;19(2):78–9.
28. Mavis BE, Brocato JJ. Postal surveys versus electronic mail surveys. The tortoise and the hare revisited. *Eval Health Prof*. 1998;21(3):395–408.
29. McMahon SR, Iwamoto M, Massoudi MS, Yusuf HR, Stevenson JM, David F, et al. Comparison of e-mail, fax, and postal surveys of pediatricians. *Pediatrics*. 2003;111(4 Pt 1):e299–303.
30. McPeake J, Bateson M, O’Neill A. Electronic surveys: how to maximise success. *Nurse Res*. 2014;21(3):24–6.
31. Nakash RA, Hutton JL, Jørstad-Stein EC, Gates S, Lamb SE. Maximising response to postal questionnaires—a systematic review of randomised trials in health research. *BMC Med Res Methodol*. 2006;6:5.
32. Pappas N, Lawrence JT, Donegan D, Ganley T, Flynn JM. Intraobserver and interobserver agreement in the measurement of displaced humeral medial epicondyle fractures in children. *J Bone Joint Surg Am*. 2010;92(2):322–7.
33. Passmore C, Dobbie AE, Parchman M, Tysinger J. Guidelines for constructing a survey. *Fam Med*. 2002;34(4):281–6.
34. Penson DF, Wei JT. *Clinical research methods for surgeons*. 1st ed. Totowa: Humana Press; 2006.
35. Pomerantz ML, Glaser D, Doan J, Kumar S, Edmonds EW. Three-dimensional biplanar radiography as a new means of accessing femoral version: a comparative study of EOS three-dimensional radiography versus computed tomography. *Skelet Radiol*. 2015;44(2):255–60.
36. Richards BS, Sucato DJ, Konigsberg DE, Ouellet JA. Comparison of reliability between the Lenke and King classification systems for adolescent idiopathic scoliosis using radiographs that were not premeasured. *Spine (Phila Pa 1976)*. 2003;28(11):1148–56; discussion 1156–7.
37. Roberts LM, Wilson S, Roalfe A, Bridge P. A randomised controlled trial to determine the effect on response of including a lottery incentive in health surveys [ISRCTN32203485]. *BMC Health Serv Res*. 2004;4(1):30.
38. Rubenfeld GD. Surveys: an introduction. *Respir Care*. 2004;49(10):1181–5.
39. Schleyer TK, Forrest JL. Methods for the design and administration of web-based surveys. *J Am Med Inform Assoc*. 2000;7(4):416–25.
40. Shiono PH, Klebanoff MA. The effect of two mailing strategies on the response to a survey of physicians. *Am J Epidemiol*. 1991;134(5):539–42.
41. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86(2):420–8.
42. Sierles FS. How to do research with self-administered surveys. *Acad Psychiatry*. 2003;27(2):104–13.

43. Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Phys Ther.* 2005;85(3):257–68.
44. Sprague S, Quigley L, Bhandari M. Survey design in orthopaedic surgery: getting surgeons to respond. *J Bone Joint Surg Am.* 2009;91(Suppl 3):27–34.
45. Sudman S. Applied sampling. In: Rossi PH, Wright JD, Anderson AB, editors. *Handbook of survey research.* San Diego: Elsevier; 1983. p. 145–94.
46. Thelen P, Delin C, Folinais D, Radier C. Evaluation of a new low-dose biplanar system to assess lower-limb alignment in 3D: a phantom study. *Skelet Radiol.* 2012;41(10):1287–93.
47. VanDenKerkhof EG, Parlow JL, Goldstein DH, Milne B. In Canada, anesthesiologists are less likely to respond to an electronic, compared to a paper questionnaire. *Can J Anaesth.* 2004;51(5):449–54.
48. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med.* 1998;17(1):101–10.
49. Wright RW, MARS Group. Osteoarthritis classification scales: interobserver reliability and arthroscopic correlation. *J Bone Joint Surg Am.* 2014;96(14):1145–51.
50. Zelnio RN. Data collection techniques: mail questionnaires. *Am J Hosp Pharm.* 1980;37(8):1113–9.



R. Kyle Martin, Andreas Persson, Håvard Visnes,
and Lars Engebretsen

Clinical Vignette: Part I

Sally is a 17-year-old female who has sustained an isolated injury to her non-dominant knee and presents for a surgical consultation with Dr. Hansen. Examination and imaging reveal an isolated complete tear of her anterior cruciate ligament (ACL). She is an elite handball player who hopes to return to her sport in the future. Her diagnosis is explained, and surgical reconstruction of her ACL is recommended.

Sally responds by saying, “I had a feeling it might be my ACL. I spoke with two of my friends who had their ACLs reconstructed; one had her hamstring used as a graft and the other used her patella tendon. Which one do you recommend? Is there a difference in failure rate?”

This chapter will explore the expanding role of registries in orthopaedic surgery with a focus on why these databases are so important. It will include examples from the Norwegian Knee Ligament Register and tips on how to maximize the usefulness of a register.

R. K. Martin (✉) · L. Engebretsen
Division of Orthopedics, Department of Orthopaedic Surgery, Oslo University Hospital, Oslo, Norway

Faculty of Medicine, University of Oslo,
Oslo, Norway

Oslo Sports Trauma Research Center, The Norwegian School of Sport Sciences, Oslo, Norway
e-mail: ummart36@myumanitoba.ca

A. Persson
Division of Orthopedics, Department of Orthopaedic Surgery, Oslo University Hospital, Oslo, Norway

The Norwegian Knee Ligament Registry, Orthopedic Department, Haukeland University Hospital, Bergen, Norway

39.1 Background/Introduction

Information from registry data can be used to guide and improve patient care and help answer questions like these that are commonly seen in clinical practice [16]. In medicine, a registry is a

H. Visnes
Division of Orthopedics, Department of Orthopaedic Surgery, Oslo University Hospital, Oslo, Norway

Oslo Sports Trauma Research Center, The Norwegian School of Sport Sciences, Oslo, Norway

The Norwegian Knee Ligament Registry, Orthopedic Department, Haukeland University Hospital, Bergen, Norway

Department of Orthopaedic Surgery, Haukeland University Hospital, Bergen, Norway

standardized database prospectively collecting information on a patient population with a common disease or intervention and followed over time [6].

The first known medical registry was the National Leprosy Registry of Norway which was established in 1856 [14, 15]. More than a century later, the Swedish Knee Arthroplasty Registry became the first national registry in orthopaedic surgery in 1975 [17]. Finland (1980), Norway (1987), and Denmark (1995) also created arthroplasty registers and were soon expanded to include all joint replacements. The primary objective of the first arthroplasty registers was the early detection of inferior results based on implant revision data. This proved successful in 1995 when two studies identified inferior implants at an early stage, a determination made possible through the arthroplasty registry [8, 11, 12].

Today, the use of registries in orthopaedic surgery has expanded to include many subspecialties other than arthroplasty, and several national, regional, and local registries have been developed. Building on the early experience of the arthroplasty registers, important changes have also been made to the data collection and outcome measures that are captured. The first knee ligament surgery registry was created in Norway in 2004, followed in 2005 by Sweden and Denmark [9]. While revision surgery and conversion to total knee arthroplasty were determined to be important outcome measures, it was also recognized that inferior results and graft failures may not always go on to further surgery and therefore may go undetected by the registry [8]. To account for this, the knee ligament registries also include patient-reported outcome measures (PROMs), specifically the Knee Injury and Osteoarthritis Outcome Score (KOOS) [28], preoperatively and at standard postoperative time points. Thus, the ability to detect inferior results and early failures is improved without relying on subsequent surgery as the sole end point.

In 2014 the International Society of Arthroplasty Registries created a PROM Working

Group [26, 27]. The working group outlined the rationale for the inclusion of PROMs in the arthroplasty registries and noted that several have incorporated PROMs into their data collection [27]. They also made several recommendations regarding the use of these outcome measures in arthroplasty registers [26].

Fact Box 39.1

Drolet and Johnson developed a definition of a medical registry based on the presence of five characteristics which they called MDR-OK. The information must be mergeable (M) into a centralized data set, use standardized data (D), and must follow a protocol outlining rules (R) for systematic and prospective data collection. Additionally, the observation (O) of patient data is followed over time to collect knowledge (K) about patient outcomes and results. The authors state that all five characteristics need to be present in order to differentiate a medical data registry from a non-registry database [6].

Clinical Vignette: In the Operating Room

Dr. Ng is just beginning a standard ACL reconstruction in a 26-year-old soccer player. The initial injury occurred 6 months ago, and the clinical evaluation and MRI both suggested an isolated ACL rupture. In the preoperative waiting room, the patient stated that he had a giving-way recently with immediate joint swelling. Upon insertion of the arthroscope, Dr. Ng finds a grade four cartilage lesion which likely developed during the recent instability episode. He has previously managed similar lesions with microfracture treatment, but he recalls a recent registry study that found microfracture at the time of ACL reconstruction

was associated with worse patient-reported outcomes than debridement [29]. He therefore chooses to change his clinical practice and performs a debridement of the lesion and then continues on with ACL reconstruction.

39.2 Importance of Registries

The goal of orthopaedic registries is to improve healthcare delivery via prospective surveillance of surgical outcomes. Continuous feedback to hospitals and surgeons allows comparison to national averages and can identify best clinical practices [8]. This, in turn, encourages improvement by setting and continually updating the standard of care [19, 20]. The detection of procedures and devices that result in early failure can be identified by following revision or reoperation rates and deterioration in patient-reported outcome measures [8, 31]. Prognostic variables associated with good and poor outcomes can be ascertained via large cohort studies performed on the data contained within the registry [8]. Epidemiological trends and the burden of disease can also be followed for changes over time.

One of the biggest advantages of national registries is the ability to include a high volume of data over time. This creates a large database of short- and long-term follow-ups from which cohort studies can be undertaken. This confers several important strengths: (1) there is little or no selection bias to influence these large data sets; (2) the data already exists at the onset of a study, making the analysis less time-consuming and more cost-effective; (3) if subsequent surgery on included patients can be linked with a personal health identification number, there is no risk for attrition bias; (4) data is collected independently of future research questions—there is no differential misclassification; and (5) the high volume of cases and data collected also allows

the study of several end points and exposures at the same time.

While randomized control trials (RCTs) are considered the gold standard research design for evaluating and comparing interventions, they are often not practical or possible to perform owing to ethical, financial, procedural, or other barriers. RCTs also often have stringent inclusion and exclusion criteria which limit the external validity and the generalizability of the results to patients seen in clinical practice. In contrast, well-designed observational studies from registry data can offer similar results regarding treatment effects while largely avoiding the hurdles faced by RCTs. These studies can be complementary to RCTs by assessing real-world applicability of the experimental findings and may serve to generate new ideas for future RCTs [1, 4, 5, 13, 22, 24].

Registries in orthopaedic surgery also have their limitations. Nonrandomized cohort studies are subject to bias from confounding variables which must be corrected for, either by selection of homogeneous subgroups or through multiple regression analysis [10]. Even when possible risk factors are considered in a regression model, there is always a risk of unmeasured confounding in observational studies due to variables that are either not considered in the model or not yet recognized or collected. Compliance is essential for an effective registry database and can be difficult to achieve, especially in the initial stages if no prior database existed. Further, the need for a high-response rate can conflict with the goal of optimizing the amount of useful data that is to be collected. To encourage form completion by the surgeons and patients, details regarding demographics, diagnosis, surgical procedure, and implants, along with objective and subjective outcome measures, must be streamlined to avoid survey fatigue and non-compliance. Finding the optimal balance between details collected and the surgeon's precious time can prove to be difficult. Consequently, the relevance of every requested input has to be carefully considered by the steering board members to keep compliance high.

Clinical Vignette: The Surgeon's Perspective

Dr. Zhang has just left the operating room following a routine ACL reconstruction with quadrupled hamstring tendon autograft. With him, he has the barcode stickers from the femoral suspension fixation device and for the interference screw he used on the tibia. He sits down with these stickers and begins filling out the one-page National Knee Ligament Registry form (Fig. 39.1). The stickers are affixed to the corresponding locations on the back of the form. The completed form will be forwarded to the central registry database, and the written patient consent and a copy of the form will remain archived locally at the patient's hospital. The entire process took less than 2 min, and Dr. Zhang is ready to see his next patient.

39.3 Research Questions

When creating a registry, it is important to understand the research questions that can be answered through the database as this will dictate what information should be recorded. Generally speaking, registries can be used to track epidemiology and provide quality assessment of treatment protocols, devices, and outcomes for a defined patient population. Regional variations within a country can also be identified.

The Norwegian Knee Ligament Registry (NKLK) provides an example of how this information can be used to provide quality control. Epidemiological data gives an actual number of ACL reconstructions and revision surgeries performed on an annual basis. Using this data, the NKLK established that a failure of a specific device is suspected if only 14 patients with this device are identified as having failed, based on recorded outcome measures [8]. This early warning system is ongoing, and problems may be identified long before they would have been uncovered by traditional methods such as RCTs. While causality of failure may not always be evident in these cases, it raises flags and directs further assessment and research.

Selecting which outcome measures and variables to record has been eluded to above. Revision surgery and consequent conversion to arthroplasty are two end points that are clear and indisputable. In addition, patient-reported outcome measures offer a subjective end point useful in identifying inferior results that may not proceed to further surgery. Selection of patient-based subjective outcome measures to include should take several variables into consideration. These include validation for use given the target population, availability in multiple languages, cost, and completion should be self-explanatory and fast (ideally less than 10 min) [8, 26].

Data recorded by the surgeon should be minimal and necessary, generally limited to a one-page reporting system. As it can be beneficial to compare data between registries, it is also important to use a core minimum data set that is in use across several different registries. In the development of the NKLK, this data was chosen based on the following three criteria [8]:

1. Can the question addressed be clearly specified and justified?
2. Is the question clinically relevant?
3. Can the item be completed postoperatively while dictating the surgery notes, not needing to seek information from other sources?

Finally, as medical practice is dynamic and ever-changing, variables should be re-evaluated on a regular basis and changes made according to current practice and evolving literature.

Fact Box 39.2

Avoid the desire to collect as much information as possible. An inverse relationship between the amount of information requested and the quality of the data obtained has been demonstrated. Completeness and accuracy of the data are where the value lies, so do not make the data set exhaustive [25]. Incomplete data is often useless data.

NATIONAL KNEE LIGAMENT REGISTRY
 Norwegian Arthroplasty Register
 Department of Orthopedic Surgery
 Haukeland University Hospital
 Mellandalsbakken 11
 N-5021 BERGEN, NORWAY
 Tlf: (+47) 55976450

Patient ID and date of birth (11 digits).....
 Name.....
 Hospital.....

CRUCIATE LIGAMENTS

CRUCIATE LIGAMENT SURGERY AND ALL REVISIONS on patients with previous cruciate ligament surgery.
 All stickers (except patient ID) are pasted in predefined columns on the back of the form.

INDEX SIDE (mark one) (Bilateral surgery= 2 forms)
 Right Left

OPPOSITE KNEE Normal Previous ACL/PCL-injury

PREVIOUS SURGERY IN INDEX KNEE (one or more)
 ACL MCL PLC Medial meniscus
 PCL LCL Cartilage Lateral meniscus
 Other, specify

DATE OF INJURY (mm.yy) | | | | | |

ACTIVITY THAT LEAD TO INJURY
 Soccer Martial arts Work
 Handball Basketball Traffic
 Alpine skiing Cross country skiing Volleyball
 Snowboard Recreational activities Skateboard
 Icehockey/bandy/
 inline skating Outdoor life Trampoline
 Racket sports Other recreational Dance
 activities
 Other.....

ACTUAL INJURY (Register all injuries – independent of surgery)
 ACL MCL PLC Menisci
 PCL LCL Cartilage
 Other.....

FURTHER INJURIES (none, one or more)
 Vasular Nerve Fracture
 Rupture in extensor apparatus
 Specify:
 N. tibialis N. peroneus
 Femur Tibia Fibula
 Patella Not sure
 Quadriceps tendon
 Patellar tendon

DATE OF SURGERY (dd.mm.yy) | | | | | |

ACTUAL SURGERY (mark one)
 (If none, skip to the next question)
 Reconstruction of cruciate ligament Revision

OTHER PROCEDURES (none, one or more)
 Meniscus surgery Osteosynthesis
 Synovectomy Cartilage surgery
 Mobilizing in narcosis Arthroscopic débridement
 Remove implant Surgery due to infection
 Bone resection (Notchplasty) Bone transplantation
 Osteotomy Arthrodesis
 Other

CHOICE OF GRAFT (see back for instructions)

	ACL	PCL	MCL	LCL	PLC
<input type="checkbox"/> BPTB					
<input type="checkbox"/> ST – double					
<input type="checkbox"/> ST – quadruple					
<input type="checkbox"/> STGR – single					
<input type="checkbox"/> STGR – double					
<input type="checkbox"/> STGR – quadruple					
<input type="checkbox"/> BQT					
<input type="checkbox"/> BQT-A					
<input type="checkbox"/> BPTB-A					
<input type="checkbox"/> BACH-A					
<input type="checkbox"/> Suture					
<input type="checkbox"/> Synthetic graft					
<input type="checkbox"/> Other					

FIXATION DEVICES
 Paste stickers in predefined columns on the back of the form
 Differentiate between femur and tibia

ACTUAL TREATMENT OF MENISCAL LESION

	Resection	Suture	Synthetic fixation*	Meniscus Transplant.	Tre-panation	None
Med.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Lat.	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

* Paste stickers in predefined columns on the back of the form

CARTILAGE LESION (none, one or more. Remember to fill in the area)
 Injury: new old undefined

	Size		ICRS Grade* (1-4)	Probable cause** (1-5)	Treatment code*** (1-9)
	Area (cm²)	≤2			
Patella MF	<input type="checkbox"/>	<input type="checkbox"/>			
Patella LF	<input type="checkbox"/>	<input type="checkbox"/>			
Trochlea fem.	<input type="checkbox"/>	<input type="checkbox"/>			
Med. fem. cond.	<input type="checkbox"/>	<input type="checkbox"/>			
Med. tib. plat.	<input type="checkbox"/>	<input type="checkbox"/>			
Lat. fem. cond.	<input type="checkbox"/>	<input type="checkbox"/>			
Lat. tib. plat.	<input type="checkbox"/>	<input type="checkbox"/>			

*ICRS Grade: 1 Nearly normal: Superficial lesions, soft indentation and/or superficial fissures and cracks; 2 Abnormal: Lesions extending down to <50% of cartilage depth; 3 Severely abnormal: Cartilage defects extending down >50% of cartilage depth as well as down to calcified layer; 4 Severely abnormal: Osteochondral injuries, lesions extending just through the subchondral boneplate or deeper defects down into trabecular bone.
 ** Probable cause: 1 Trauma; 2 CM: chondromalacia patellae; 3 OCD: osteochondritis dissecans; 4 OA: primary osteoarthritis; 5 Other: Specify cause in correct column
 *** Treatment code: 1 Debridement; 2 Microfracture; 3 Mosaic; 4 Biopsy for cultivation; 5 Cell transplantation; 6 Cell transplantation with matrix; 7 Periosteum transplantation; 8 No treatment; 9 Other: Specify cause in correct column

OUTPATIENT SURGERY No Yes

PER OPERATIVE COMPLICATIONS No Yes, which.....

DURATION OF SURGERY (skin to skin-time).....min.

SYSTEMIC ANTIBIOTIC PROPHYLAXIS No Yes

Name (A)
 Dosage (A)..... Total number of dosagesDurationhours
 Dosage (B)..... Total number of dosagesDurationhours

TROMBOSIS PROPHYLAXIS

No
 Yes, name
 Dosage Duration.....days
 First dosage given preoperative No Yes
 If second prophylaxis is used:
 Dosage Duration.....days
 Stocking No Calf Thigh Duration.....days
 Other, specify

Surgeon:.....
 Surgeon that filled in the form (surgeon's name is not registered).

Fig. 39.1 Norwegian National Knee Ligament Registry Perioperative surgeon registration form

Clinical Vignette: The Patient Perspective

Robert is a 23-year-old student who ruptures his ACL playing recreational basketball and is referred to see an orthopaedic surgeon. His diagnosis and options are discussed, and he agrees to proceed with ACL reconstruction. On the day of surgery, he is given two forms for completion in the preoperative holding area. The first is an informed consent form for inclusion in the knee ligament registry. It includes information about the NKLR, the type of information recorded, data protection, and the follow-up procedure and informs Robert that he may be invited to participate in future research projects. He also learns that he may withdraw from the NKLR at any time and that confidentiality is ensured. The second form is a Knee Injury and Osteoarthritis Outcome Score form, which he completes in approximately 10 min.

After surgery, Robert progresses through his rehabilitation and experiences a regular postoperative course. At 2, 5, and 10 years after surgery, he receives a KOOS form in the mail or e-mail for completion. Three additional questions are included at the end of the form inquiring (1) whether he has sustained any new injuries or had further surgery on the operated knee; (2) if so, what was injured; and (3) if he re-ruptured his ACL, how was the diagnosis made (MRI, arthroscopy, physician examination, or examination by other healthcare professional). If Robert forgets to submit his KOOS, he will receive a reminder after 3 months.

is obtained from the patient and from the surgeon at prespecified time points. In general, baseline subjective data is obtained from the patient preoperatively and repeated at defined intervals postoperatively. Information related to the diagnosis, surgical findings, procedure(s), implants, and complications is recorded by the surgeon immediately following the surgery [8] (Fig. 39.2). Barcode stickers from implanted devices can often be scanned or included with the report form to ensure accurate tracking [8, 10]. Some registries also include clinical and/or radiological follow-up data at subsequent postoperative visits [18, 21].

All documents are sent to a central location to be checked for completeness and stored in the registry database. Incomplete forms are returned to the sender for completion to assure data quality and completeness. A copy of the form is also retained in the patient's hospital chart. In Norway, a small staff is responsible for the day-to-day operations of the NKLR, overseen by an advisory board. This includes a secretary, a computer engineer, and an administrative head of the registry. Each participating hospital employs secretarial assistance, and the NKLR also has access to experienced statisticians for registry studies [8]. In 2017, the operating budget for the central NKLR office was 1,800,000 NOK (approximately \$218,000 USD) and includes the salary of additional staff involved in NKLR-based research projects.

Orthopaedic registries are most often publicly financed either through direct funding or through grants. The Norwegian, Swedish, and Danish knee ligament registries are financed through the national public health system. Similarly, the Canadian Joint Replacement Registry receives funding from federal and provincial government sources, via the Canadian Institutes for Health Information [3]. Privately owned and financed registries also exist, for example, in New Zealand. Collaboration with the local and national orthopaedic association is important in all cases to advocate for continued financing and support ongoing activities of the registry. In general, it is advisable for registries to avoid industry sponsorship to maintain objectivity and avoid bias which may affect research and clinical practice.

39.4 Structure of a Registry

The registry collects prospective information on all patients in a defined population. Patient consent to participate may or may not be required, based on the specific national privacy legislation. Information

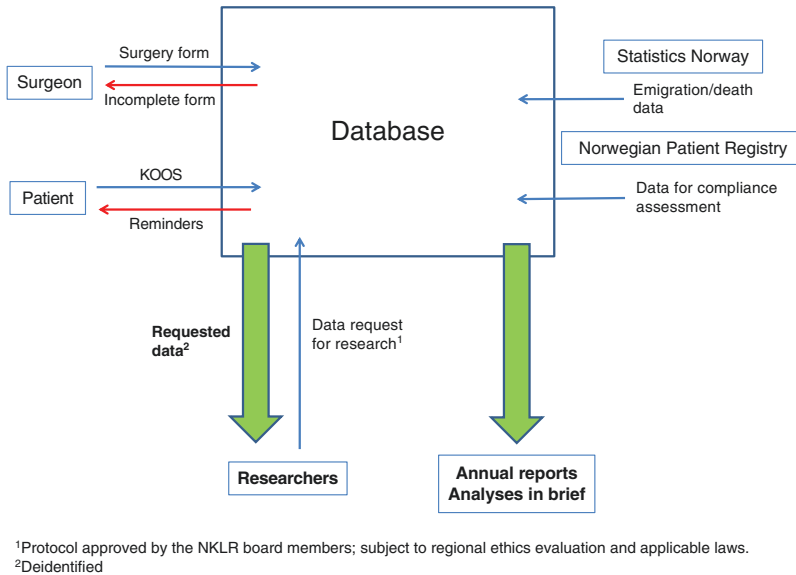


Fig. 39.2 Norwegian National Knee Ligament Registry flow diagram

39.5 Data Reporting

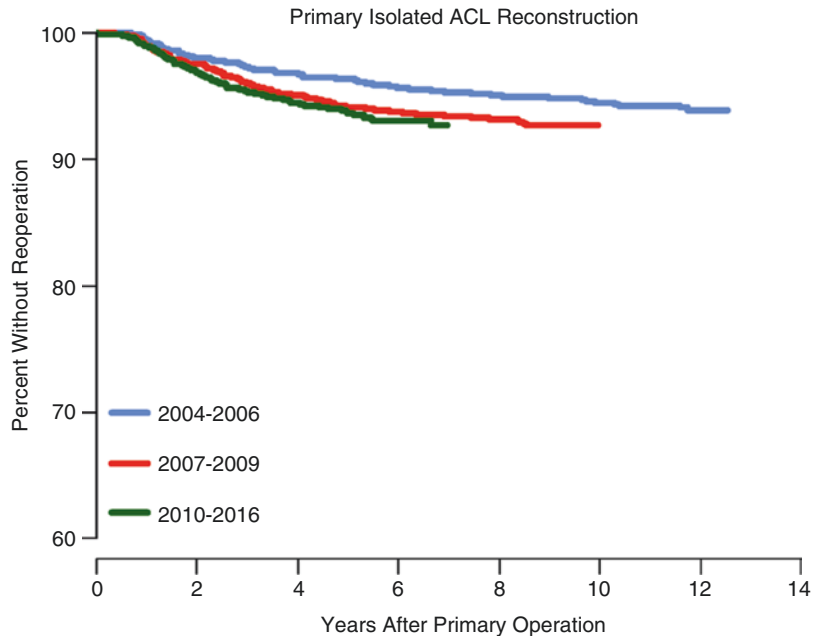
Data contained in the registry can be released for evaluation via annual reports, hospital-specific data requests, or more extensive research projects. Annual reports present descriptive information including compliance rates and are often broken down into regional subsets for comparison to the national and historical data. Tables and figures are used to present overall and regional epidemiology, survivorship curves, complications, and other information (Fig. 39.3). These reports are frequently published online for public viewing.

Since attaining and maintaining a high compliance rate are paramount to the success of a registry, it follows that accurate determination of this rate is of equal importance. Hospital, regional, or national databases that capture information related to the number of procedures performed are often used as the denominator in determining the compliance rate of registries. For example, in Norway, hospitals track their procedures and send the information to the central office for reimbursement according to the diagnosis-related groups schedule. This data is

then compared with the information received by the registry over the same time period to determine the compliance rate. Any discrepancies are included in the annual report as hospital-specific compliance rates.

Registries may report only national and regional data, keeping surgeons unidentified, or they may use the information to produce surgeon-specific reports [8]. In one Canadian province, annual surgeon-specific reports are generated for all hip and knee arthroplasties, hip fractures, rotator cuff repairs, and knee meniscectomies. These report cards allow hospitals and individual surgeons to compare their data with that of their peers. Further, the regional Standards and Quality Committee also reviews the reports and notifies surgeons that fall below the regional averages. If outcomes remain below the average for 2 or more years, the committee meets with the surgeon to review the data and to develop a plan to correct the identified issues. This initiative is possible owing to the mandatory nature of regional registry participation, strong leadership of the steering group, and legislation protecting the reports from court subpoena [2, 30]. Concerns over surgeon confidentiality, legal ramifications, and the poten-

Fig. 39.3 Kaplan-Meier survivorship curve of primary ACL reconstruction from the Norwegian National Knee Ligament Registry 2017 annual report



tial effect on compliance rate limit the effectiveness of this model in other jurisdictions. One must also be aware that using end points in a register as an indicator of healthcare quality might be influenced by differences in patient populations between hospitals/clinics that might affect the outcome measure. For example, surgical centres that perform more complex procedures or treat high-risk patients may have outcomes that differ from the national average.

Access to information in addition to the annual report is also encouraged. Requests for hospital-specific data often require approval by the advisory board and are made by the official hospital contact person. This potentially patient- and surgeon-identifiable information can be used by the hospital for evaluation and local quality improvement projects. More extensive requests for information to be used in research projects can be applied for in writing to the governing body, and additional Health Research Ethics Board approval may also be required [8].

Clinical Vignette: Obtaining Data for a Study

Dr. Perrin wishes to compare the revision rates between ACL reconstructions performed using autograft and allograft over the past 5 years, using the national registry data. To request this data, a written application is made to the steering committee. In it, Dr. Perrin includes:

1. A complete project proposal, including the name of the principle investigator who bears responsibility for the project
2. A description of the problem
3. The data selection requested and the variables to be used in the analysis
4. The publication plan and timeline

Dr. Perrin also submits an application to the local Health Research Ethics Board. The steering committee reviews the appli-

cation and determines if the purpose of the study is aligned with the purpose of the national registry. The application is further evaluated on the professional relevance of the project, the capacity of the investigator to produce a satisfactory analysis using the data, and with consideration of other ongoing projects that may be similar. Once satisfied, approval is granted and the data is released for the study.

39.6 Pearls

The strength of a registry lies in the completeness and accuracy of the data [25], and therefore compliance is crucial. This can be difficult to achieve for several reasons that have been described previously. Some ways to improve compliance include the way in which data is collected, rules and laws governing its collection, and perception regarding the importance and uses of the registry.

Whether paper-based or electronic forms are used, they should be user-friendly and not time-consuming. Only the necessary data should be included, and if possible, as much of the data should be recorded in advance (i.e. operating room nurses can enter the operative information and scan the implant barcodes during the surgery [30]). Missing data should be flagged immediately and a notice sent to the surgeon to address the deficiency before too much time has passed after the surgery.

The collection of personal information is and always should be an important discussion. Several jurisdictions require informed consent prior to enrollment in a registry, while others do not. Clearly this can play a role in compliance, with one national registry reporting that 31% of the submitted forms were missing the corresponding consent [3]. Strict adherence to confidentiality standards including secure data storage and limited authorized access may bridge the gap between privacy concerns and data collection. Decisions

related to the necessity of obtaining consent must ultimately involve regional Health Research Ethics Boards and established legislation.

Legislation can influence compliance with national registries in other ways. In Denmark, the hospitals do not receive reimbursement for cruciate ligament surgeries that have not been reported to their national knee ligament registry. Creating a user-friendly data submission environment for the surgeon and introducing policy mandating compliance is a powerful way to maintain a high rate of data completion on a national level.

Perception of the registry in the eyes of the public and the orthopaedic community also plays a role in compliance. This begins on the first encounter with each patient which should include a discussion on the importance of the registry and an overview of their contribution. It is also an opportunity to build a good rapport that may influence their future compliance. The regular publication of reports highlighting trends and important findings, both positive and negative, provides constant reminder that the data is being collected for a purpose. Major publications that may change clinical practice further reinforce the importance of participation in the registry. Finally, data should be used to foster best medical practice, rather than seeking to identify and punish individual surgeons who may fall below this current standard. A supportive community approach to the registry should be sought to ensure that all surgeons feel comfortable participating and remain open to the feedback it provides.

39.7 Future Directions

Moving forward, registries will continue to play an important role in orthopaedic surgery. As more national registries are established, there is a vision to create a common international knee ligament registry in Europe. The development of a common software program to collect and store the data is an additional goal that could be used

by those nations who will not join an international database for legal or other reasons. Standardizing data collection across several nations would have several benefits including increased power for large studies and the ability to directly compare data from one part of the world with another. Finally, there is ongoing effort to expand registries to include the nonoperative management of orthopaedic conditions such as ACL deficiency which are not currently captured in most databases [7].

Clinical Vignette: Part II

Sally is still in Dr. Hansen's office and he is reflecting on her questions. Regarding the differences between hamstring tendon and patella tendon autografts, he discusses the pros and cons of each with respect to donor site morbidity, anterior knee pain, skin incisions, and complications including infection rates. He advises her that either one is a perfectly acceptable option but also mentions a recent review of the Norwegian Knee Ligament Registry suggesting a revision rate of hamstring tendons that was twice that of patella tendon grafts [23]. As Dr. Hansen is a surgeon in Norway, his patients were included in that cohort study. Since its publication he now recommends patella tendon autograft to patients like Sally while still presenting all options. Sally is satisfied and agrees to proceed with ACL reconstruction using a patella tendon autograft.

39.8 Useful/Inexpensive Resources

1. National Institutes of Health: <https://www.nih.gov/health-information/nih-clinical-research-trials-you/list-registries>
2. Norwegian National Advisory Unit on Arthroplasty and Hip Fractures: <http://nrlweb.ihelse.net/eng/>

Take-Home Message

- Registries are an indispensable research tool in orthopaedic surgery that can improve patient care and lead to significant changes in clinical practice.
- A thorough understanding of the strengths and limitations of these databases is important for those involved in registry-based study design and the interpretation of results.
- In the future, increased international collaboration, focus on patient-reported outcome measures, and the inclusion of non-surgical management strategies will further enhance the value of these registries.

References

1. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med.* 2000;342:1878–86. <https://doi.org/10.1056/NEJM200006223422506>.
2. Bohm ER. Personal communication with Dr. E. Bohm, Professor of Surgery, University of Manitoba; Chair, Canadian Joint Replacement Registry Advisory Committee; Chair, Manitoba Provincial Orthopaedic Standards and Quality Committee. 2017.
3. Bohm ER, Dunbar MJ, Bourne R. The Canadian Joint Replacement Registry—what have we learned? *Acta Orthop.* 2010;81:119–21. <https://doi.org/10.3109/17453671003685467>.
4. Comber H, Perry JJ. Observational studies for intervention assessment. *Lancet.* 2001;357:2141–2. [https://doi.org/10.1016/S0140-6736\(00\)05219-3](https://doi.org/10.1016/S0140-6736(00)05219-3).
5. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med.* 2000;342:1887–92. <https://doi.org/10.1056/NEJM200006223422507>.
6. Drolet BC, Johnson KB. Categorizing the world of registries. *J Biomed Inform.* 2008;41:1009–20. <https://doi.org/10.1016/j.jbi.2008.01.009>.
7. Engebretsen L, Forssblad M, Lind M. Why registries analysing cruciate ligament surgery are important. *Br J Sports Med.* 2015;49:636–8. <https://doi.org/10.1136/bjsports-2014-094484>.
8. Granan L-P, Bahr R, Steindal K, Furnes O, Engebretsen L. Development of a national cruciate ligament surgery registry: the Norwegian National Knee Ligament Registry. *Am J Sports Med.* 2008;36:308–15. <https://doi.org/10.1177/0363546507308939>.
9. Granan L-P, Forssblad M, Lind M, Engebretsen L. The Scandinavian ACL registries 2004–2007: baseline epidemiology. *Acta Orthop.* 2009;80:563–7. <https://doi.org/10.3109/17453670903350107>.
10. Havelin LI, Engesaeter LB, Espehaug B, Furnes O, Lie SA, Vollset SE. The Norwegian Arthroplasty

- Register: 11 years and 73,000 arthroplasties. *Acta Orthop Scand.* 2000;71:337–53. <https://doi.org/10.1080/000164700317393321>.
11. Havelin LI, Espehaug B, Vollset SE, Engesaeter LB. The effect of the type of cement on early revision of Charnley total hip prostheses. A review of eight thousand five hundred and seventy-nine primary arthroplasties from the Norwegian Arthroplasty Register. *J Bone Joint Surg Am.* 1995;77:1543–50.
 12. Havelin LI, Espehaug B, Vollset SE, Engesaeter LB. Early aseptic loosening of uncemented femoral components in primary total hip replacement. A review based on the Norwegian Arthroplasty Register. *J Bone Joint Surg Br.* 1995;77:11–7.
 13. Horan FT. Judging the evidence. *J Bone Joint Surg Br.* 2005;87:1589–90. <https://doi.org/10.1302/0301-620X.87B12.17247>.
 14. Irgens LM. The origin of registry-based medical research and care. *Acta Neurol Scand.* 2012;126:4–6. <https://doi.org/10.1111/ane.12021>.
 15. Irgens LM, Bjerkedal T. Epidemiology of leprosy in Norway: the history of The National Leprosy Registry of Norway from 1856 until today. *Int J Epidemiol.* 1973;2:81–9.
 16. Ivers N, Jamtvedt G, Flottorp S, Young JM, Odgaard-Jensen J, French SD, O'Brien MA, Johansen M, Grimshaw J, Oxman AD. Audit and feedback: effects on professional practice and healthcare outcomes. In: *The Cochrane Collaboration, editor. Cochrane Database of Systematic Reviews.* Chichester: Wiley; 2012.
 17. Knutson K, Lewold S, Robertsson O, Lidgren L. The Swedish knee arthroplasty register. A nation-wide study of 30,003 knees 1976–1992. *Acta Orthop Scand.* 1994;65:375–86.
 18. Lind M, Menhert F, Pedersen AB. The first results from the Danish ACL reconstruction registry: epidemiologic and 2 year follow-up results from 5,818 knee ligament reconstructions. *Knee Surg Sports Traumatol Arthrosc.* 2009;17:117–24. <https://doi.org/10.1007/s00167-008-0654-3>.
 19. Malchau H, Herberts P, Eisler T, Garellick G, Söderman P. The Swedish total hip replacement register. *J Bone Joint Surg Am.* 2002;84-A(Suppl 2):2–20.
 20. Maloney WJ. National Joint Replacement Registries: has the time come? *J Bone Joint Surg Am.* 2001;83-A:1582–5.
 21. Mygind-Klavsen B, Grønbech Nielsen T, Maagaard N, Kraemer O, Hölmich P, Winge S, Lund B, Lind M. Danish Hip Arthroscopy Registry: an epidemiologic and perioperative description of the first 2000 procedures. *J Hip Preserv Surg.* 2016;3:138–45. <https://doi.org/10.1093/jhps/hnw004>.
 22. Naylor CD, Guyatt GH. Users' guides to the medical literature. X. How to use an article reporting variations in the outcomes of health services. The Evidence-Based Medicine Working Group. *JAMA.* 1996;275:554–8.
 23. Persson A, Fjeldsgaard K, Gjertsen J-E, Kjellens AB, Engebretsen L, Hole RM, Fevang JM. Increased risk of revision with hamstring tendon grafts compared with patellar tendon grafts after anterior cruciate ligament reconstruction: a study of 12,643 patients from the Norwegian Cruciate Ligament Registry, 2004–2012. *Am J Sports Med.* 2014;42:285–91. <https://doi.org/10.1177/0363546513511419>.
 24. Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? *N Engl J Med.* 2000;342:1907–9. <https://doi.org/10.1056/NEJM200006223422511>.
 25. Robertsson O. Knee arthroplasty registers. *J Bone Joint Surg Br.* 2007;89:1–4. <https://doi.org/10.1302/0301-620X.89B1.18327>.
 26. Rolfson O, Bohm E, Franklin P, Lyman S, Denissen G, Dawson J, Dunn J, Eresian Chenok K, Dunbar M, Overgaard S, Garellick G, Lübbecke A, Patient-Reported Outcome Measures Working Group of the International Society of Arthroplasty Registries. Patient-reported outcome measures in arthroplasty registries: Report of the Patient-Reported Outcome Measures Working Group of the International Society of Arthroplasty Registries Part II. Recommendations for selection, administration, and analysis. *Acta Orthop.* 2016;87:9–23. <https://doi.org/10.1080/17453674.2016.1181816>.
 27. Rolfson O, Eresian Chenok K, Bohm E, Lübbecke A, Denissen G, Dunn J, Lyman S, Franklin P, Dunbar M, Overgaard S, Garellick G, Dawson J, The Patient-Reported Outcome Measures Working Group of the International Society of Arthroplasty Registries. Patient-reported outcome measures in arthroplasty registries: Report of the Patient-Reported Outcome Measures Working Group of the International Society of Arthroplasty Registries Part I. Overview and rationale for patient-reported outcome measures. *Acta Orthop.* 2016;87:3–8. <https://doi.org/10.1080/17453674.2016.1181815>.
 28. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. *J Orthop Sports Phys Ther.* 1998;28:88–96. <https://doi.org/10.2519/jospt.1998.28.2.88>.
 29. Røtterud JH, Sivertsen EA, Forssblad M, Engebretsen L, Årøen A. Effect on patient-reported outcomes of debridement or microfracture of concomitant full-thickness cartilage lesions in anterior cruciate ligament-reconstructed knees: a nationwide cohort study from Norway and Sweden of 357 patients with 2-year follow-up. *Am J Sports Med.* 2016;44:337–44. <https://doi.org/10.1177/0363546515617468>.
 30. Singh J, Politis A, Loucks L, Hedden DR, Bohm ER. Trends in revision hip and knee arthroplasty observations after implementation of a regional joint replacement registry. *Can J Surg.* 2016;59:304–10. <https://doi.org/10.1503/cjs.002916>.
 31. de Steiger RN, Miller LN, Davidson DC, Ryan P, Graves SE. Joint registry approach for identification of outlier prostheses. *Acta Orthop.* 2013;84:348–52. <https://doi.org/10.3109/17453674.2013.831320>.

Part VIII

How to Perform an Economic Health Care Study?



How to Perform an Economic Healthcare Study

40

Jonathan Edgington, Xander Kerman, Lewis Shi,
and Jason L. Koh

40.1 Introduction

Research in orthopedic surgery focuses on finding treatments and diagnostic tests that provide the best outcomes for patients, but the concept of what is best for the patient is constantly evolving. The focus has broadened from objective outcome measures to include patients' subjective appreciation of those outcomes and is expanding to include the costs of the interventions as well. Rising healthcare costs are an impetus to study how healthcare dollars are spent with a goal of delivering high-quality healthcare [3, 7], especially as rapid advances in orthopedic technology with hope of improved outcomes create questions of whether advances provide improved value. For these reasons, it will become increasingly important for orthopedic surgeons to understand how to appropriately perform economic healthcare stud-

ies. This chapter reviews the methods of performing economic healthcare studies and relates them to clinical cases in orthopedic surgery.

40.1.1 What Is Healthcare Economics?

Economics is a social science that studies production and distribution of goods and services in the context of scarcity. Broadly, there are two types of economics: positive economics and normative economics [7]. Positive economics is objective and primarily concerned with evaluating the present (e.g., the United States spends \$8.2 billion on orthopedic surgery annually) [14], while normative economics is subjective, inherently prescriptive, and forward-looking (e.g., the United States should spend less than \$8.2 billion on orthopedic surgery annually). This distinction is important as healthcare economic studies often focus on positive economics in order to objectively study the relationships between costs, outcomes, and distributions in the delivery of healthcare. These subjects are important for assessing the efficiency, value, and equity of medical care and can then be used to make informed normative economic conclusions.

J. Edgington (✉) · L. Shi
Department of Orthopaedic Surgery and
Rehabilitation Medicine, University of Chicago
Medical Center, Chicago, IL, USA
e-mail: Jonathan.Edgington@uchospitals.edu;
lshi@bsd.uchicago.edu

X. Kerman
Pritzker School of Medicine, The University
of Chicago, Chicago, IL, USA
e-mail: Alexander.Kerman@uchospitals.edu

J. L. Koh
NorthShore Orthopaedic Institute, NorthShore
University HealthSystem, Evanston, IL, USA

40.1.2 What Do Efficiency, Value, and Equity Mean in Healthcare Economics?

Several core economic definitions are relevant to healthcare economics. Technical efficiency means that no additional output can be achieved given a certain amount of inputs (i.e., no surgical materials wasted). Production efficiency means that inputs have been optimized for a given level of production (i.e., operating rooms consistently at capacity but not over/under-booked). Allocative efficiency means that resources have been distributed to those who will benefit the most from those resources, thus maximizing utility (i.e., patients who most need surgery receive surgery first). Generally, the term efficiency in healthcare economics refers to the cost of care to achieve a certain measure of quality.

Value is a closely related but crucially different concept from efficiency, because it measures a stakeholder's subjective appreciation of an outcome. This stakeholder is generally the patient, but value can also be considered from the perspectives of payors, providers, or society as a whole. A common definition of value in healthcare is $\text{Value} = (\text{Outcomes})/(\text{Costs})$ [7, 9].

Equity concerns the allocative efficiency of distribution of finite resources among a group of people. Equity may not line up with efficiency in which case these conflicting concepts must be reconciled with the help of economic healthcare studies [12]. These definitions are further summarized in Fact Box 40.1.

40.1.3 What Are the Goals of Healthcare Economic Studies in Orthopedics?

Orthopedic care often focuses on increasing quality of life or increasing patient mobility through improving function of the musculoskeletal system. These tasks have proved costly, with Medicare estimates of orthopedic expenditures surpassing \$8.2 billion annually [14]. These high costs provide an incentive for clinicians to prove the value of their services.

Fact Box 40.1: Healthcare Economic Terms Defined

Technical efficiency: producing maximal output from the minimal quantity of inputs
Production efficiency: a situation when production goods cannot be increased without decreasing production of another good
Allocative efficiency: the point at which marginal cost of production and additional good equals marginal benefit of the good
Value: the perceived benefit of a good or service
Equity: resource distribution that is fair among a group

Economic studies in orthopedics should not limit themselves to simply identifying the cheapest intervention, but should rather focus on understanding the costs of achieving improved outcomes for patients, payors, and society by combining clinical and economic data. With value data in hand, physicians can provide better care for their patients.

40.2 Research Question

As in any research project, the first step of an economic orthopedic study is to identify a study question and develop a hypothesis. The problem or question often arises in one of two ways: as follow-up on previous published work or as original concepts following an observation. Before initiating a study, the concepts in the following section should be explored to establish a clear overview of the study plan.

40.2.1 What Is the Question?

The research question is the foundation of all parts of the study, and it is worth putting additional time and effort into crystallizing before beginning work. The question may arise out of personal interest, gaps in literature, or translation of topics from other fields. Often deep thought about the topic leads to a plethora of further questions or ambiguity about the purpose of the study. When

dealing with potentially ambiguous topics like costs and benefits in economic healthcare studies, it is important to clearly define the question in order to guide both planning and execution.

40.2.2 What Is the State of Current Knowledge on the Topic?

Knowing what other researchers have learned on the topic is important in order to prevent wasted effort duplicating results and to further focus the research question. While mainstream medical journals increasingly publish healthcare economic studies, economic- and quality improvement-focused publications often feature these types of projects as well. Finding literature that addresses the same question does not preclude the investigator from performing a study but is useful for comparison and learning from previous shortcomings. Review of current literature will focus the question and justify prioritizing this question.

40.2.3 Why Is the Question Important?

Having a solid understanding of the current literature and knowledge helps define the deficiencies in evidence and establish the importance of the question. This will add purpose and meaning to the study. Answering the question *why would it be helpful to know the answer to this question?* establishes how the study will help future clinicians, researchers, and policymakers.

40.2.4 Hypothesis

With this information the clinician can form an explicit and a concise hypothesis. The hypothesis

should be a theory or prediction of what the outcome of the study may show. The hypothesis should be established well before data is collected or analyzed, meaning researchers should avoid collecting data and then retrospectively developing the hypothesis. It should be obvious how the hypothesis will be tested given the circumstances of the study. The hypothesis should not only be valuable if found to be true but also should be valuable if found to be false. With a clear research question and hypothesis, one can proceed with appropriate selection of an economic evaluation for a given study.

40.3 Economic Evaluation and Application to Research Question

Once a research question and a hypothesis have been identified, the next step is identifying the appropriate economic analytic method (Table 40.1). As seen in Table 40.2, there are four common analytic approaches in healthcare economics, each with advantages and disadvantages: cost-minimization (CMA), cost-effectiveness (CEA), cost-utility (CUA), and cost-benefit (CBA) analyses. All four rely on accurately quantifying costs but vary in their considerations of outcomes. Deciding on which best suits the study is based on the research question, hypothesis, and purpose of the study, as well as the data available.

Cost may be quantified in a variety of ways depending on the perspective of interest. Common perspectives include total costs to society (societal costs), costs to patients (patient costs), and costs to third-party payors (payor costs). Direct costs may measure medical or non-medical costs associated with treatment. Other

Table 40.1 Potential results following cost-effectiveness analysis

	More effective	Less effective
Higher Cost	Consider this option	Do not choose this intervention
Lower cost	Choose this intervention	Consider this option

Table 40.2 Multiple sources of possible financial and cost data with examples in parenthesis

Source of cost or financial data
Individual patient records
Practice records
Hospital records
Healthcare system records (Kaiser Permanente)
Healthcare consortium databases (e.g., the National Surgical Quality Improvement Project, Vizient)
State or regional databases (New York State, California)
National government databases (e.g., in the United States, the National Inpatient Survey)
Private insurance databases (PearlDiver)

Fact Box 40.2: Methods of Performing Economic Analysis

	Outcomes	Cost
Cost-minimization analysis (CMA)	Assumed equivalent	Dependent on perspective of interest (patient, payor, societal)
Cost-effectiveness analysis (CEA)	Natural units (objective, quantity, OR quality)	
Cost-utility analysis (CUA)	Utility scale (subjective + objective, quantity, AND quality)	
Cost-benefit analysis (CBA)	Monetary scale (subjective + objective, quantity, AND quality)	

intangible costs, such as opportunity costs, may be considered in the quantification depending on what is pertinent to the research question [7]. For example, when a patient elects to undergo surgery, the time spent recovering cannot be spent earning income, and therefore lost wage income is an opportunity cost associated with the surgery.

Outcomes may be quantified in a variety of ways as well. Objective clinical outcomes can be defined very broadly based on the research topic, ranging from functional benefits, improved outcome scores, to morbidity. Subjective clinical outcomes can also be measured in a variety of ways using validated instruments to assess patients’ perspectives of their health (e.g., patient-reported outcome tools). Objective and subjective outcomes can also be combined together in a variety of ways. One commonly used combined measure is the quality-adjusted life year (QALY), which provides a standardized score based on not only remaining quantity of life but also the quality, or health state, of those years. A variety of instruments exist to broadly measure health state on a 0 to 1 scale with 0 being death and 1 being perfect health. Another form of combined subjective and objective outcome

measurement is the monetary value that patients would be willing to pay to achieve different outcomes and costs can be seen in Fact Box 40.2.

40.3.1 Cost-Minimization Analysis

Cost-minimization analysis (CMA) is used when the evaluation targets a difference in cost of interventions with the assumption that outcomes are equivalent. In order for this technique to be valid, equivalent outcomes must be clearly demonstrated comprehensively. The advantage of a cost-minimization analysis is simplicity. The disadvantage of CMA is establishing complete equivalency is often difficult or impossible. Even if equivalent outcomes were felt to be true, there is always the possibility that equivalence may actually be a failure to detect a difference in two interventions. If outcomes cannot be proved equivalent, cost-minimization analysis is not an appropriate economic evaluation technique. Some economists argue that CMA techniques are very seldom appropriate, with possible exceptions in rare cases where randomized control trials have already demonstrated equivalent outcomes [2]. CMA are most frequently relevant

for policy or third-party payor decisions on resource allocations where concerns of specific outcomes come secondary to cost considerations. If one assumed the outcomes of hyaluronic acid injections and corticosteroid injections were equally effective, one may estimate cost difference with a CMA evaluation.

40.3.2 Cost-Effectiveness Analysis

Cost-effectiveness analysis (CEA) is a method used to compare the costs and outcomes of one intervention versus the costs and outcomes of an alternative. This type of evaluation provides a value of additional units of cost per additional unit of benefit. The set of outcomes must be measured with the same units of effectiveness between both interventions. As seen in Table 40.3, this leads to four possible outcomes. The advantage of CEA is the ability to directly compare costs and outcomes of two interventions. The disadvantage of CEA is only one outcome may be measured even if the intervention may have multiple relevant outcomes. For example, a total hip arthroplasty may provide measurable outcomes such as relief of pain, increased range of motion, and decreased narcotic usage, but only one can be considered at a time in CEA. An excellent example of this type of evaluation is demonstrated by Rajan et al. (2018) in their study on surgical fixation of distal radius fractures [10]. Because of this short coming, CEA does not fully provide an answer to the value question: *Is the additional collective benefit gained worth the extra cost?*

40.3.3 Cost-Utility Analysis

Cost-utility analysis (CUA) differs from CEA in that CUA outcomes measured using a health

Table 40.3 An example of how comparator choice may impact conclusions

	Cost	Time	Accuracy (%)
Osteotome and mallet	\$500	2 min	80
Hand saw	\$25	5 min	95
Hand rasp	\$25	4 h	99.9

index that attempts to quantify all health outcomes associated with an intervention. Most often quality-adjusted life years (QALY) is the index of choice which attempts to objectify length of life as well as quality of life related to the intervention. A variety of instruments exist to calculate QALYs by broadly estimating health quality states on a 0 to 1 scale using a multi-attribute utility model (MAU, 0 = no utility of life/death and 1 = perfect health) and multiplying this number by the life expectancy benefit of a procedure: $QALY = (MAU) * (LEB)$. Several MAUs are commonly used including the EuroQoL-5 Dimensions, Health Utilities Index, Short Form-6 Dimensions, Assessment of Quality of Life, 15 Dimensions, and Quality of Well-Being [13]. This type of analysis is highly appropriate in orthopedics, as research in this field typically looks at increased quality of life rather than simply increasing length of life. CUA using QALYs from the societal perspective is recommended by the Panel on Cost-Effectiveness in Health and Medicine convened by the US Public Health Service in 1996 and 2016 [15]. CUA has been effectively used to evaluate QALY following total hip arthroplasty using QALYs [3–5].

40.3.4 Cost-Benefit Analysis

Cost-benefit analysis (CBA) is an analysis that places a monetary value on both the cost of the intervention and the outcomes of the intervention. This allows for an analysis that establishes a net benefit of an intervention. This evaluation attempts to comprehensively consider what patients value in their healthcare. The advantage of this analysis is you have an objective net benefit which can answer the question *is the additional benefit gained worth the additional cost?* The disadvantage of this type of evaluation is assigning monetary values to clinical outcomes is often difficult and sometimes controversial, requiring social science research expertise and validation. An example of CBA in orthopedics would be assigning dollar values to the net benefits gained following total hip arthroplasty, which

would allow a patient-centric conclusion about the value of THA.

40.4 Perspective

The perspective of the study is the lens through which the study is viewed and drives the assessments of cost and benefit. Relevant perspectives in healthcare economic studies include the patients', providers', hospitals', insurers', or a comprehensive societal perspective. The perspective can be thought of as the target audience for any conclusions drawn from the study. It is important as the priorities of the different groups likely differ significantly. Changing the perspective changes both the numerator and denominator of the value equation because different elements must be included depending on whose perspective is under consideration. For example, if an evaluation is executed from an insurer's perspective, the definition of cost would likely exclude patients' out-of-pocket expenses. In certain instances, it may be valuable to include multiple perspectives in each study, in which case results of evaluations should be reported separately from each perspective.

40.5 Challenges

Healthcare economic studies present several types of challenges not frequently encountered in other projects. These challenges include difficulty defining and acquiring data on costs and benefits, as well as the underlying issue of what interventions are compared to one another. The conclusions reached in healthcare economic studies also may be difficult to apply in practice or be limited in their generalizability to other procedures or practice settings.

40.5.1 Defining Costs

Financial or cost data can be obtained from multiple sources, from the individual practice level to national databases. An example of the variability

Table 40.4 Steps to performing an economic healthcare study

Identify the relevant question
Understand current knowledge
Formulate hypothesis
Obtain relevant financial information
Obtain relevant outcome information
Perform appropriate statistical analysis
Report results
Discuss results with respect to question of interest

in calculating costs depending on the methodology used can be found in Palsis et al. (2018) [8]. Several examples of cost or financial data are listed in Table 40.4. In many cases, it can be difficult or expensive to obtain cost or financial data.

For investigators, identifying an accurate source of information is critical. Typically, data that is derived more directly from clinical records is more accurate than that derived from administrative databases, which can be subject to errors in coding. Nevertheless, valuable information can still be obtained from administrative databases.

Additionally, data that is obtained exclusively from one clinical site (e.g., a hospital) may not accurately reflect the total costs of care, including items such as costs related to postoperative rehabilitation. As financial analysis of healthcare-related costs becomes more sophisticated, there is increasing ability to more fully understand costs. In addition, allocation of costs can be done in multiple ways: they can be related to charges, known costs (such as for implants), or activity-based costing.

Finally, the economic data must be tied to clinical outcomes for patients. Assessment of clinical outcomes is a critically important area and is further discussed in other chapters. Relating economic and outcome data should be performed with appropriate statistical rigor.

40.5.2 Defining Outcomes

Compared to defining costs, outcomes may be even more challenging to capture completely. Outcome data is often gathered from clinical research and prior literature. The most reliable

benefit or outcome information would be from a double-blind randomized control trial (RCT). Few orthopedic RCT comparing surgical outcomes exist, and therefore much of the orthopedic outcomes data is observational [1, 6, 11]. Observational studies can be pooled or meta-analyses can be used; however these introduce bias or error. Outcome data should be as accurate and applicable as possible, and as above, methods of obtaining the data should be explained in detail.

40.5.3 Comparator Choice

Choosing the correct interventions to compare to one another presents further challenges in healthcare economic studies. Appropriate comparator choice helps provide correct conclusions following analysis. For example, consider tools for making cuts during a total knee arthroplasty as seen in Table 40.3. The hand saw is more effective and cheaper than an osteotome and mallet, so the lattermost can immediately be rejected. The decision then is whether the additional 4.9% accuracy is worth the additional 3 h and 55 min of time spent. If one were to compare only an osteotome and mallet combination to a hand rasp, the erroneous conclusion of the hand rasp as the best tool will be made. Choosing appropriate comparators can be quite difficult in orthopedics and vastly impacts conclusions drawn following analysis.

40.5.4 Clinical Applicability

As many economic evaluations may take on the broad perspectives of society or a hospital system, clinical applicability of the conclusions of these studies may be difficult. Translation of “macro” results may be difficult into the perspectives of the provider or patient. Consider a study executed from a societal perspective that shows hyaluronic acid injections are not cost-effective compared to cortisone injections for knee arthritis. Cost may seem irrelevant to a patient who is considering trying another type of injection versus undergoing a major surgery such as a total

knee replacement. Studies may have limited clinical applicability in cases where cost differences are statistically different but not clinically significant. For example, suppose a study showed cost savings of \$50 per patient for anterior versus posterior hip arthroplasty. This may not be significant to the patient or payors if total costs are in the \$10,000–\$20,000 range.

40.5.5 Generalizability

Many economic evaluations rely on multiple assumptions and estimations that may not be universally true. Accepting assumptions and estimations allows the study to provide results and conclusions. However, this shortcoming means many evaluations cannot be generalized to other settings or populations. The vast difference in cost for services throughout the healthcare field makes generalization statements extremely difficult.

Take-Home Message

- Given the rising healthcare costs, the role of economic healthcare studies in orthopedics will certainly increase over the coming years. Just as evidence-based medicine brought about changes in clinical decision-making in years past, economic healthcare studies will likely present themselves as necessary knowledge.
- We suggest that researchers follow a step-by-step progression in performing an economic healthcare study as seen in Table 40.4.
- This chapter has provided a brief overview for the practicing orthopedic surgeon who may be interested in performing such study or understanding results of these types of studies.

References

1. Brauer CA, Rosen AB, Olchanski NV, Neumann PJ. Cost-utility analyses in orthopaedic surgery. *J Bone Joint Surg Am.* 2005;87:1253–9. <https://doi.org/10.2106/JBJS.D.02152>.
2. Briggs AH, O'Brien BJ. The death of cost-minimization analysis? *Health Econ.* 2001;10:179–84. <https://doi.org/10.1002/hec.584>.

3. Chang RW, Pellisier JM, Hazen GB. A cost-effectiveness analysis of total hip arthroplasty for osteoarthritis of the hip. *JAMA*. 1996;275:858–65.
4. Daigle ME, Weinstein AM, Katz JN, Losina E. The cost-effectiveness of total joint arthroplasty: a systematic review of published literature. *Best Pract Res Clin Rheumatol*. 2012;26:649–58. <https://doi.org/10.1016/j.berh.2012.07.013>.
5. Givon U, Ginsberg GM, Horoszowski H, Shemer J. Cost-utility analysis of total hip arthroplasties. Technology assessment of surgical procedures by mailed questionnaires. *Int J Technol Assess Health Care*. 1998;14:735–42.
6. Maniadas N, Gray A. Health economics and orthopaedics. *J Bone Joint Surg Br*. 2000;82:2–8.
7. Napper M, Newland J. Health economics information resources: a self-study course: U.S. National Library of Medicine; 2016.
8. Palsis JA, Brehmer TS, Pellegrini VD, Drew JM, Sachs BL. The cost of joint replacement: comparing two approaches to evaluating costs of total hip and knee arthroplasty. *J Bone Joint Surg Am*. 2018;100:326–33. <https://doi.org/10.2106/JBJS.17.00161>.
9. Porter ME. What is value in health care? *N Engl J Med*. 2010;363:2477–81. <https://doi.org/10.1056/NEJMp1011024>.
10. Rajan PV, Qudsi RA, Dyer GSM, Losina E. The cost-effectiveness of surgical fixation of distal radial fractures: a computer model-based evaluation of three operative modalities. *J Bone Joint Surg Am*. 2018;100:e13. <https://doi.org/10.2106/JBJS.17.00181>.
11. Rajan PV, Qudsi RA, Wolf LL, Losina E. Cost-effectiveness analyses in orthopaedic surgery: raising the bar. *J Bone Joint Surg Am*. 2017;99:e71. <https://doi.org/10.2106/JBJS.17.00509>.
12. Reidpath DD, Olafsdottir AE, Pokhrel S, Allotey P. The fallacy of the equity-efficiency trade off: rethinking the efficient health system. *BMC Public Health*. 2012;12:S3. <https://doi.org/10.1186/1471-2458-12-S1-S3>.
13. Wisløff T, Hagen G, Hamidi V, Movik E, Klemp M, Olsen JA. Estimating QALY gains in applied studies: a review of cost-utility analyses published in 2010. *Pharmacoeconomics*. 2014;32:367–75. <https://doi.org/10.1007/s40273-014-0136-z>.
14. Trends in orthopaedics: an analysis of medicare claims, 2000–2010. <https://www.healio.com/orthopaedics/journals/ortho/2013-3-36-3/%7Bed4e1047-4f2f-46bc-9f01-65f36a2efd58%7D/trends-in-orthopaedics-an-analysis-of-medicare-claims-20002010>. Accessed 13 Feb 2018.
15. Recommendations from the second panel on cost-effectiveness in health and medicine | Guidelines | JAMA | The JAMA Network. <https://jamanetwork.com/journals/jama/article-abstract/2552214?redirect=true>. Accessed 15 Feb 2018.

Part IX

Multi-Center Study: How to Pull It Off?



Conducting Multicenter Cohort Studies: Lessons from MOON

41

José F. Vega and Kurt P. Spindler

41.1 Introduction

The Multicenter Orthopaedic Outcomes Network (MOON) is the largest prospective longitudinal anterior cruciate ligament (ACL) reconstruction cohort with at least 80% follow-up in the United States. MOON is a collaboration of 17 surgeons from 7 different large academic medical centers (Cleveland Clinic Foundation, Vanderbilt Orthopaedic Institute, The Ohio State University, University of Iowa, Washington University in St. Louis, Hospital for Special Surgery, and University of Colorado) that has collected >80% follow-up at 2, 6, and 10 years post-op of more than 4400 ACL reconstructions [5, 13, 19, 27]. To date, the MOON cohort has produced over 40 peer-reviewed publications; possibly more importantly, it has created a template for conducting high-quality, prospective orthopedic research that has been utilized by other large, high-quality, prospective, multicenter studies [4, 18, 19, 34]. This chapter chronicles the planning, development, and execution of MOON.

J. F. Vega
Cleveland Clinic Lerner College of Medicine,
Cleveland, OH, USA
e-mail: vegaj@ccf.org

K. P. Spindler (✉)
Cleveland Clinic Sports Health Center,
Garfield Heights, OH, USA
e-mail: spindlk@ccf.org, stojsab@ccf.org

41.2 The Early Years: Envisioning MOON and Laying the Foundation

The first MOON patient was enrolled in January of 2002, but the story of MOON actually begins a decade earlier, in the early 1990s, with a much smaller prospective cohort.

At the time (the late 1980s and early 1990s), it was clear that ACL repair was an ineffective strategy for treating ACL rupture, as failure rates were high and patients did poorly, especially after the first and second postoperative years [8, 21]. It was also established that chronic ACL deficiency led to poor outcomes and the early-onset posttraumatic osteoarthritis [1, 17]. However, new and encouraging data suggested that “augmented” ACL reconstruction (primary repair with the assimilation of a bone-patellar tendon-bone [BTB] autograft into the repair) led to improved knee function and outcomes [7].

Thus, the senior author (KPS) assembled a cohort of 54 patients that underwent acute (within 3 months of injury) ACL reconstruction. Patients were enrolled by the senior author, who was completing a sports medicine fellowship at the Cleveland Clinic at the time. This cohort was assembled with the intention to follow the participants for 10 years in order to better understand the importance of bone bruising and articular cartilage lesions on ACL reconstruction outcomes [11, 30].

However, it did not take long to appreciate that the outcome of an ACL reconstruction is influenced by a myriad of factors beyond bone bruising and articular cartilage lesions, and, as a result, 54 patients would not be enough. This was also a period of transition, as the use of primary ACL repair was nearly abandoned, while the utilization of ACL reconstruction was growing in popularity. Thus, new questions were appearing at a rate much greater than we could answer.

As the sun began to rise high in the midmorning sky, the senior author and one of his Cleveland Clinic colleagues (JTA) found themselves cycling through the elevation changes of Sun Valley, Idaho, mulling over potential ways to address the seemingly endless list of questions surrounding ACL reconstruction.

It was then that they admitted, reluctantly, that the only feasible way to answer so many questions would be to assemble another cohort, this time with 10 or 20 times the number of patients they had enrolled in Cleveland. Naturally, the discussion then turned to how they could possibly enroll and follow hundreds (and more likely thousands) of patients undergoing a procedure which, at the time, was being performed less than 90,000 times per year across the entire country [3]. The answer, of course, was to develop a multicenter network.

41.2.1 The Vanderbilt Sports Medicine (VSM)-Cleveland Clinic Foundation (CCF) ACL Reconstruction Registry

Having finished his fellowship at the Cleveland Clinic, the senior author accepted a position as an assistant professor at Vanderbilt University Medical Center. This transition came with exciting potential, as two sites could certainly enroll more patients than either site alone, and, thus, the VSM-CCF ACL Reconstruction Registry was born.

Between the fall of 1991 and spring of 1998, the three surgeons (KPS, JTA, and RP) involved in the VSM-CCF ACL Reconstruction Registry enrolled a total of 1201 patients and followed them until the fifth postoperative year. Around

the same time, the VSM-CCF ACL Reconstruction Registry gained a new partner in the form of The Ohio State University, which was maintaining its own ACL reconstruction database. This three-institution cohort, containing 2286 ACL reconstructions performed over a 10-year period, yielded a total of eight publications on a wide range of topics [2, 9, 10, 15, 22, 23, 29, 31].

What made this multicenter cohort so unique was not only its size (at the time, it was one of the largest prospective ACL reconstruction cohorts in the country) but also its use of patient-reported outcome measures (PROMs) as primary endpoints. The collection of preoperative and intraoperative variables, along with the follow-up PROMs, allowed for complex multivariable regression analysis to identify clinically relevant predictors of outcomes at 5 years [31]. Although difficult to imagine given the state of today's literature, use of PROMs as a primary endpoint was tantamount to clinical research heresy in the early and mid-2000s [12, 35]. Nonetheless, the developers of this registry adopted two recently validated PROMs to measure outcomes following ACL reconstruction—the International Knee Documentation Committee Subjective Knee Form (IKDC-SKF) and the Knee injury and Osteoarthritis Outcome Score (KOOS).

Fact Box 41.1

The use of patient-reported outcome measures has exploded over the last two decades. In a retrospective review of 4 major orthopedic surgery journals, Siljander et al. found that 94 publications used patient-reported outcome measures in 2004, compared to 228 in 2016 [26].

Aside from the hesitancy of the scientific community to accept the use of PROMs as primary endpoints, the utilization of the IKDC and the KOOS created an additional problem—missing baseline data, as the KOOS was not developed until 1998, and the IKDC debuted 3 years later, in 2001 [14, 25]. As a result, this cohort was unable to capture the true impact of ACL recon-

struction (or the prognostic significance of pre- and intraoperative variables), as there was no baseline PROM data which would be needed in order to observe change over time. Thus, once again, there was a need for a new cohort—one that would be larger, comprised of patients from more centers, and from which more variables could be collected.

41.3 Designing MOON

41.3.1 Assembling the Clinical Research Team

The MOON principal investigator (PI) and co-investigators recognized that designing a prospective longitudinal cohort would require a large team comprised of individuals with clinical research expertise beyond orthopedic surgeons. They were aided by one of orthopedic surgery's original clinician-scientists—Sandy Kirkley—who had specialized training in clinical research long before it became commonplace (although still rare today, a master's in public health [MPH] or MPH equivalent was virtually unheard of in orthopedics in the 1990s). In addition, we had several experts in prospective trial and cohort design, an epidemiologist, and a PhD biostatistician with specialized training in multivariable analysis. Furthermore, they hired a program manager and several research assistants. Without this expertise, all outside of orthopedics, MOON would have certainly not been successful at both achieving NIH funding and maintaining high rates of longitudinal follow-up.

41.3.2 Aims

The first step in designing the MOON cohort was to identify which questions were to be answered by MOON. Among the many options, the co-investigators settled on three primary aims:

1. What preoperative and intraoperative variables predict both short- and long-term outcomes following ACL reconstruction?
2. What predicts graft failure?

3. What is the natural progression of posttraumatic osteoarthritis, and what variables modify the trajectory?

41.3.3 Sample Size

Due to the fact that the co-investigators wanted to assess long-term outcomes, as well as the development of osteoarthritis, this new cohort would need to be followed for at least 10 years. Since they also wanted to identify predictors of graft failure, this new cohort would need to be large. Considering that failure was seen in roughly 10% of the three-institution ACL reconstruction registry, and that multivariable logistic regression was to be done to identify significant predictors of graft failure, MOON would need to enroll at least 2250 participants (15 participants per variable included in the logistic regression, multiplied by 15 variables to be included, multiplied by 10 because failure would likely occur in only 10% of the cohort).

To enroll such a large number of patients in a reasonable period of time, the designers of MOON set out to assemble a team of surgeons and various high-volume ACL reconstruction institutions that could enroll roughly 600 patients per year. Building off of the VSM-CCF ACL reconstruction registry with the help of personal connections, the final tally of MOON surgeons and sites came to 17 surgeons at 7 institutions (Cleveland Clinic Foundation, Vanderbilt Orthopaedic Institute, The Ohio State University, University of Iowa, Washington University in St. Louis, Hospital for Special Surgery, and University of Colorado).

41.3.4 Outcomes

In the same fashion as the co-investigators' three-institutional ACL reconstruction registry, the decision was made to use PROMs as a primary endpoint. Again, they opted to utilize the IKDC and the KOOS to track PROMs from baseline until the completion of the study (tentatively projected to be 10 years post-op).

The decision to use PROMs as a primary endpoint was twofold. First, PROMs had been

demonstrated to be valid methods of measuring outcomes over time and were directly relevant to both clinicians and patients; second, the use of PROMs would allow the investigators to follow a large cohort without incurring insurmountable expense or significant losses to follow-up, as participants could complete PROMs from the comforts of their own homes.

41.3.5 Data Collection

The next challenge to overcome was to develop a data collection system that would not be overly burdensome to patients and/or surgeons. With some guidance from their late friend (SK), the designers compiled a series of paper questionnaires that captured all of the pertinent variables and trialed the Compaq iPAQ (Fig. 41.1) as a means for electronic data capture. This was no small feat, especially considering that the Apple iPhone would not debut for another 6 years.

Luckily, electronic databases such as REDCap are now available and make excellent options for collecting large amounts of data rapidly across multiple sites.

Clinical Vignette

In the early stages of MOON, the typical work flow involved completion of paper PROMs which were then faxed to a central location for data collection. As mentioned above, the iPAQ was trialed briefly, but, because of difficulties with reliable syncing and data upload, was abandoned quickly.

41.3.6 Agreement Studies

While having so many fellowship-trained surgeons was undoubtedly an advantage, managing a group as large and experienced as that was not without its own challenges. The most obvious hurdle facing the group was how to come to an agreement on how certain variables would be



Fig. 41.1 Compaq iPAQ, 2004

classified (e.g., meniscal tear location, depth, type, etc.) and then treated (e.g., meniscectomy vs. meniscal repair). The group performed a pair of inter-rater agreement studies to demonstrate, in a scientific way, that their methodology classifying each particular variable (specifically, meniscus tears and articular cartilage lesions) was reliable across the group [6, 20].

A similar challenge emerged with respect to the method in which each surgeon's technique for performing a typical ACL reconstruction. Again, the group conducted two studies in which they demonstrated that the differences between the way in which each surgeon performed his preferred ACL reconstruction were minimal and likely clinically irrelevant [32, 33]. In the first study, 12 surgeons each performed 6 ACL reconstructions on cadaveric knees using their technique of preference (e.g., two-incision, trans-tibial, or anteromedial). The 72 knees were then imaged in order to identify any significant differences in tunnel characteristics (which none were present) [33]. Four surgeons then went on to duplicate the same study design in MOON patients, complete with postoperative CT scans to assess each tunnel. Again, no differences were identified [32].

The last obstacle to overcome with regard to agreement was which graft type to use. Rather than attempt to convince all of the participating surgeons to perform their ACL reconstructions with BTB or hamstring (which would have been

impossible), the group showed, through a systematic review, that no clinically relevant difference exists between BTB and hamstring autograft [28]. In hindsight, this was actually a judicious approach, as the inclusion of both BTB and hamstring ACL reconstructions only provided further data to conclusively demonstrate that no clinically relevant difference exists between the two approaches when done correctly.

41.3.7 Funding

As expected, enrolling and following such a large cohort would come with significant cost (e.g., the estimated annual cost of operationalizing MOON at Vanderbilt University alone was nearly \$200,000).

MOON enrolled and followed patients for nearly 3 full years before submitting its first application for a National Institutes of Health (NIH) RO1 grant in February 2004 (which, by the way, was not selected for funding). By the end of 2005 (still before MOON was funded by the NIH), MOON had enrolled 2340 patients. At 2 years post-op, 93% had phone follow-up, and 85% had returned completed PROMs. Of course, achieving such high follow-up rates for a cohort as large as that came with a large price tag. MOON had four full-time employees devoted to seeking follow-up (one employee could follow roughly 600 patients per year), and, still, MOON surgeons had to personally call 25% of their patients in order to get them to follow-up. The nested cohort within MOON (aimed at understanding the development of posttraumatic osteoarthritis) also came with significant expense, which included training radiology personnel at three separate sites, specialized bilateral weight-bearing radiographs, and blinded evaluation by surgeons and physical therapists. It was not until 2006, after three submissions and three rejections, that MOON received NIH funding.

In total, MOON cost roughly \$1.4 million to operate between 2001 (when the first patient was enrolled) and 2006 (when MOON was first awarded grant funding). Of the \$1.4 million,

roughly half was provided by a combination of unrestricted gifts (Smith & Nephew [\$450,000] and Aircast [\$200,000]) and a grant from the National Football League (NFL) Charities (\$125,000). The other half was covered by an internal tax placed on one of MOON's designers (KPS) and two of his partners (ECM, JK) at Vanderbilt University (\$450,000) and the remaining money from an OREF Prospective Clinical Research Grant (\$149,000).

Fact Box 41.2

The award rate (defined as the number of awards given in a fiscal year divided by the total number of applications reviewed, including new applications and resubmissions from the previous review period) for NIH grant proposals between 1990 and 2014 fluctuated between 15 and 30%, although it has not exceeded 20% since 2005 [24].

41.4 MOON Becomes a Reality

Finally, in January of 2002, MOON was ready to enroll its first patient. But the challenges did not stop there. It was only because of teamwork and dedication that MOON was able to become a reality. New and/or ongoing concerns were discussed during monthly conference calls (a call that has taken place on the second Monday of each month for the last 17 years and continues to this day). Each site continues to maintain an active institutional review board (IRB) application.

Since it received initial funding in 2006, MOON has successfully renewed its funding on three separate occasions, allowing for 6- and 10-year patient follow-up. Currently, MOON has >4400 ACL reconstructions in its database and has achieved >80% follow-up at 2, 6, and 10 years post-op [5, 13, 27].

To date, MOON has produced more than 40 peer-reviewed publications and has laid the foundation for other multicenter orthopedic research groups [4, 18, 19, 34].

Fact Box 41.3

Many of the clinical sites that participated in MOON went on to participate in other well-known studies including the Meniscus Tear in Osteoarthritis Research (MeTeOR) Trial, the Multicenter ACL Revision Study (MARS), and an upcoming randomized controlled trial.

Take-Home Message

- The story of MOON is one of vision, perseverance, and teamwork.
- Conducting prospective multicenter orthopedic research is possible, but it requires significant buy-in from all participants—from the PI to the research coordinators and even the study participants themselves.
- When multiple investigators are involved, inter-rater agreement studies are crucial to demonstrate the validity of the data being collected.
- Arguably the greatest key to success is regular and open communication.
- Do not be afraid to dream big. In the words of Norman Peale, “Shoot for the moon.
- If you miss it, you will still land among the stars” [16].

References

1. Andersson C, Odensten M, Good L, Gillquist J. Surgical or non-surgical treatment of acute rupture of the anterior cruciate ligament. A randomized study with long-term follow-up. *J Bone Joint Surg Am.* 1989;71(7):965–74.
2. Bowers AL, Spindler KP, McCarty EC, Arrigain S. Height, weight, and BMI predict intra-articular injuries observed during ACL reconstruction: evaluation of 456 cases from a prospective ACL database. *Clin J Sport Med.* 2005;15(1):9–13.
3. Buller LT, Best MJ, Baraga MG, Kaplan LD. Trends in anterior cruciate ligament reconstruction in the United States. *Orthop J Sports Med.* 2015;3(1):2325967114563664. <https://doi.org/10.1177/2325967114563664>.
4. Dunn WR, Kuhn JE, Sanders R, et al. 2013 Neer Award: predictors of failure of nonoperative treatment of chronic, symptomatic, full-thickness rotator cuff tears. *J Shoulder Elbow Surg.* 2016;25(8):1303–11. <https://doi.org/10.1016/j.jse.2016.04.030>.
5. Dunn WR, Spindler KP, MOON Consortium. Predictors of activity level 2 years after anterior cruciate ligament reconstruction (ACLR): a Multicenter Orthopaedic Outcomes Network (MOON) ACLR cohort study. *Am J Sports Med.* 2010;38(10):2040–50. <https://doi.org/10.1177/0363546510370280>.
6. Dunn WR, Wolf BR, Amendola A, et al. Multirater agreement of arthroscopic meniscal lesions. *Am J Sports Med.* 2004;32(8):1937–40.
7. Engebretsen L, Benum P, Fasting O, Mølster A, Strand T. A prospective, randomized study of three surgical techniques for treatment of acute ruptures of the anterior cruciate ligament. *Am J Sports Med.* 1990;18(6):585–90. <https://doi.org/10.1177/036354659001800605>.
8. Feagin JA, Curl WW. Isolated tear of the anterior cruciate ligament: 5-year follow-up study. *Am J Sports Med.* 1976;4(3):95–100. <https://doi.org/10.1177/036354657600400301>.
9. Fox JA, Nedeff DD, Bach BR Jr, Spindler KP. Anterior cruciate ligament reconstruction with patellar autograft tendon. *Clin Orthop.* 2002;402:53–63.
10. Graham SM, Parker RD. Anterior cruciate ligament reconstruction using hamstring tendon grafts. *Clin Orthop.* 2002;402:64–75.
11. Hanyasiak BT, Spindler KP, Rothrock CR, et al. Twelve-year follow-up on anterior cruciate ligament reconstruction: long-term outcomes of prospectively studied osseous and articular injuries. *Am J Sports Med.* 2008;36(4):671–7. <https://doi.org/10.1177/0363546508315468>.
12. Heckman JDME-C. Are validated questionnaires valid? [Letter]. *J Bone.* 2006;88(2):446.
13. Hettrich CM, Dunn WR, Reinke EK, MOON Group, Spindler KP. The rate of subsequent surgery and predictors after anterior cruciate ligament reconstruction: two- and 6-year follow-up results from a multicenter cohort. *Am J Sports Med.* 2013;41(7):1534–40. <https://doi.org/10.1177/0363546510290051301>.
14. Irrgang JJ, Anderson AF, Boland AL, et al. Development and validation of the international knee documentation committee subjective knee form. *Am J Sports Med.* 2001;29(5):600–13. <https://doi.org/10.1177/03635465010290051301>.
15. Kaeding CC, Pedroza AD, Parker RD, Spindler KP, McCarty EC, Andrish JT. Intra-articular findings in the reconstructed multiligament-injured knee. *Arthroscopy.* 2005;21(4):424–30. <https://doi.org/10.1016/j.arthro.2004.12.012>.
16. Kandel B. Peale still positive; Words he lives by. *USA Today.* 1988;2A.
17. Kannus P, Järvinen M. Conservatively treated tears of the anterior cruciate ligament. Long-term results. *J Bone Joint Surg Am.* 1987;69(7):1007–12.
18. Katz JN, Brophy RH, Chaisson CE, et al. Surgery versus physical therapy for a meniscal tear and osteoarthritis. *N Engl J Med.* 2013;368(18):1675–84. <https://doi.org/10.1056/NEJMoa1301408>.

19. Lynch TS, Parker RD, Patel RM, et al. The impact of the Multicenter Orthopaedic Outcomes Network (MOON) research on anterior cruciate ligament reconstruction and orthopaedic practice. *J Am Acad Orthop Surg.* 2015;23(3):154–63. <https://doi.org/10.5435/JAAOS-D-14-00005>.
20. Marx RG, Connor J, Lyman S, et al. Multirater agreement of arthroscopic grading of knee articular cartilage. *Am J Sports Med.* 2005;33(11):1654–7. <https://doi.org/10.1177/0363546505275129>.
21. Odensten M, Lysholm J, Gillquist J. Suture of fresh ruptures of the anterior cruciate ligament. A 5-year follow-up. *Acta Orthop Scand.* 1984;55(3):270–2.
22. Paul JJ, Spindler KP, Andrish JT, Parker RD, Secic M, Bergfeld JA. Jumping versus nonjumping anterior cruciate ligament injuries: a comparison of pathology. *Clin J Sport Med.* 2003;13(1):1–5.
23. Piasecki DP, Spindler KP, Warren TA, Andrish JT, Parker RD. Intraarticular injuries associated with anterior cruciate ligament tear: findings at ligament reconstruction in high school and recreational athletes. An analysis of sex-based differences. *Am J Sports Med.* 2003;31(4):601–5. <https://doi.org/10.1177/03635465030310042101>.
24. Rockey S. What are the chances of getting funded? NIH Extramur Nexus. June 2015. <https://nexus.od.nih.gov/all/2015/06/29/what-are-the-chances-of-getting-funded/>. Accessed Feb 1 2018.
25. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynnon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS)—development of a self-administered outcome measure. *J Orthop Sports Phys Ther.* 1998;28(2):88–96. <https://doi.org/10.2519/jospt.1998.28.2.88>.
26. Siljander MP, McQuivey K, Fahs A, Galasso L, Serdahely K, Karadsheh MS. Current trends in patient reported outcome measures in total joint arthroplasty: a study of four major orthopedic journals. *J Arthroplasty.* 2018. <https://doi.org/10.1016/j.arth.2018.06.034>.
27. Spindler KP, Huston LJ. O'Donoghue sports injury award 10 year outcomes and risk factors after ACL reconstruction: a multicenter cohort study. *Orthop J Sports Med.* 2017;5(7 Suppl 6):2325967117S00247. <https://doi.org/10.1177/2325967117S00247>.
28. Spindler KP, Kuhn JE, Freedman KB, Matthews CE, Dittus RS, Harrell FE. Anterior cruciate ligament reconstruction autograft choice: bone-tendon-bone versus hamstring: does it really matter? A systematic review. *Am J Sports Med.* 2004;32(8):1986–95.
29. Spindler KP, McCarty EC, Warren TA, Devin C, Connor JT. Prospective comparison of arthroscopic medial meniscal repair technique: inside-out suture versus entirely arthroscopic arrows. *Am J Sports Med.* 2003;31(6):929–34. <https://doi.org/10.1177/03635465030310063101>.
30. Spindler KP, Schils JP, Bergfeld JA, et al. Prospective study of osseous, articular, and meniscal lesions in recent anterior cruciate ligament tears by magnetic resonance imaging and arthroscopy. *Am J Sports Med.* 1993;21(4):551–7. <https://doi.org/10.1177/036354659302100412>.
31. Spindler KP, Warren TA, Callison JC, Secic M, Fleisch SB, Wright RW. Clinical outcome at a minimum of five years after reconstruction of the anterior cruciate ligament. *J Bone Joint Surg Am.* 2005;87(8):1673–9. <https://doi.org/10.2106/JBJS.D.01842>.
32. Wolf BR, Ramme AJ, Britton CL, Amendola A, MOON Knee Group. Anterior cruciate ligament tunnel placement. *J Knee Surg.* 2014;27(4):309–17. <https://doi.org/10.1055/s-0033-1364101>.
33. Wolf BR, Ramme AJ, Wright RW, et al. Variability in ACL tunnel placement: observational clinical study of surgeon ACL tunnel variability. *Am J Sports Med.* 2013;41(6):1265–73. <https://doi.org/10.1177/0363546513483271>.
34. Wright RW, Huston LJ, Haas AK, et al. Effect of graft choice on the outcome of revision anterior cruciate ligament reconstruction in the Multicenter ACL Revision Study (MARS) Cohort. *Am J Sports Med.* 2014;42(10):2301–10. <https://doi.org/10.1177/0363546514549005>.
35. Zarins B. Are validated questionnaires valid? *J Bone Joint Surg Am.* 2005;87(8):1671–2. <https://doi.org/10.2106/JBJS.E.00554>.



MARS: The Why and How of It

42

Rick W. Wright, Amanda K. Haas,
and Laura J. Huston

42.1 Why MARS

Anterior cruciate ligament (ACL) reconstruction remains the treatment of choice for ACL-deficient active individuals involved with sports or activities that involve quick start/stop, cutting, jumping, and abrupt change of direction activities. Based on industry and implant evidence, there are approximately 400–500,000 ACL reconstructions per year in the United States. Fortunately, primary reconstructions typically do well, but can fail at a low but significant rate [8, 11, 14]. While highly successful in the short term, there can be problems with primary reconstructions including loss of motion, extensor dysfunction with certain grafts, arthritis, and graft failure. This treatise will address graft failure and the multicenter, multi-surgeon group assembled to study the issues surrounding revision ACL reconstruction.

A variety of studies have evaluated graft failure in the primary ACL reconstruction setting and have found the rates to range from ~1 to 8% in the standard patient and graft setting. In the

Multicenter Orthopaedic Outcomes Network (MOON), the graft failure rate in primary reconstructions was 3% in the ipsilateral knee and 3% in the contralateral knee at 2-year follow-up [11]. In a systematic review evaluating hamstring vs. patellar tendon autografts, Spindler et al. reported a 3.7% overall failure rate (95% CI: 1.5–5.7%) [8]. At minimum 5-year follow-up, Wright et al. found in a systematic review of ipsilateral vs. contralateral failure a rate of 5.8% in the ipsilateral and 11.8% in the contralateral knee [14].

These are reasonably low levels of failure, but the question remains what happens to those that unfortunately do fail. These patients may have inappropriate expectations of their current and future knee health. They may desire to return once again to the activities that resulted in ACL failure and not realize the potential for poor results. In their mind it is like changing the oil or tires and you're good to go again. A better understanding of revision results would allow us to better counsel patients as to expected outcomes.

What outcomes can patients expect and what outcomes truly matter? This became the impetus for our current Multicenter ACL Revision Study (MARS) Group. There was consensus among these surgeons that revision ACL reconstruction typically resulted in worse outcomes compared to primary reconstructions. Revision ACL reconstruction was the strongest predictor for worse KOOS scores in a mixed ACL reconstruction

R. W. Wright (✉) · A. K. Haas
Department of Orthopaedic Surgery, Washington
University School of Medicine, St. Louis, MO, USA
e-mail: rwright@wustl.edu; ahaas22@wustl.edu

L. J. Huston
Vanderbilt Orthopaedic Institute, Nashville, TN, USA
e-mail: laura.huston@Vanderbilt.Edu

cohort early in the MOON experience. In a series of ACL reconstructions reported at minimum 5-year follow-up, revision was the strongest predictor for worse outcome across the board, but in the editing process the journal reviewers and editor requested that the revisions be removed from the study [9].

Unfortunately, little Level 1 or 2 evidence existed to help us confirm this discrepancy in outcomes between primary and revision ACL-reconstructed patients. In a mixed model meta-analysis of 21 studies with minimum 2-year follow-up after revision reconstruction, these worse results were demonstrated [13]. Of the 21 studies, however, only 4 were Level 1 or 2, while 1 was Level 3, and 16 were Level 4 studies. Objective failure (defined as a re-revision, KT-1000 >5 mm, or a positive pivot shift) occurred in $13.7 \pm 2.7\%$ —much higher than typical primary failure rates. Patient-reported outcomes were worse than expected compared with primary ACL results and usually exceeded the known clinically important difference for these outcome scores.

Given these findings, the Multicenter Orthopaedic Outcomes Network (MOON) Group reviewed their prospectively collected cohort which began as a mixed primary and revision patient ACL reconstruction cohort [10, 12]. One working hypothesis for the study was that revision ACL reconstruction results in worse outcome compared to primary reconstruction, as measured by validated patient-based outcome measures including the Marx activity level, Knee injury and Osteoarthritis Outcome Score (KOOS), and International Knee Documentation Committee Subjective form (IKDC). 487 ACL reconstructions met the following inclusion/exclusion criteria: (1) all meniscal/chondral treatment included and (2) osteotomy, posterior cruciate ligament (PCL), or collateral ligament surgery excluded. 408/487 (84%) were available at minimum 2-year follow-up with 39/47 (83%) revisions available at 2-year follow-up. At 2 years, median Marx scores had dropped from 12 to 9 points in the primary reconstructions vs. 10 to 6 points in revisions ($p = 0.009$). While the minimally clinically important difference (MCID) is not known for the Marx

score, it is assumed on a 16-point scale that 2 points (representing more than a 10% change or difference) would be clinically significant. The IKDC at 2 years was 85.6 for primary ACL reconstructions and 79.6 for revisions, which was statistically different ($p = 0.005$), but not clinically significant (MCID: 11.5 points). For the 5 KOOS subscales, a difference of 8–10 points is clinically significant. In this study, the KOOS Knee-related Quality of Life (KRQOL) subscale was lower in revisions (75 vs. 62.5) at 2 years ($p < 0.001$). The KOOS Sports and Recreation subscale also was worse for revisions (85 vs. 75; $p = 0.004$) at 2 years. KOOS Pain was lower in revisions and potentially clinically significant (91.7 vs. 83.3). KOOS Symptoms (85.7 vs. 78.6) and ADLs (98.7 vs. 97.1) did not demonstrate a clinically significant difference between primary and revision reconstructions.

Thus, this represented a prospectively collected cohort of primary and revision ACL reconstructions evaluated identically by validated patient-reported outcome measures with worse scores across the board for revisions, but with no obvious factors contributing to these worse outcomes. Also, revisions represented only 10% of the cohort. To perform multivariable analysis requires ~10–15 subjects per variable assessed, and with the multiple factors (50–75 or more) potentially contributing to revision ACL reconstruction outcomes, it would require quick assimilation of 750–1000 patients. It became apparent the MOON Group, with less than 20 members, could not enroll an adequate number of patients quickly enough for this type of study. As we jokingly say, “a simple moon could not get it done, we needed a planet.” With this in mind, we set out to establish a larger group of interested sports medicine surgeons.

We felt the basic approach utilized by the MOON Group would be appropriate for evaluating the revision patient but realized there were different factors involved in revision surgeries that would need to be captured. A small group developed a standard operating procedure (SOP) manual that outlined rules of engagement for surgeons and patients. We very early on engaged the American Orthopaedic Society for Sports

Medicine (AOSSM), through the late Bart Mann, PhD, the research director. He, the AOSSM Research Committee, and the AOSSM society were excited about the chance to offer the research opportunity to their members. Based on this we utilized their website and email to advertise to members to participate. Once interested members were identified, we set up three meetings to educate surgeons and describe how the study would proceed. We also engaged these attendees in designing forms and determining the variables we would collect. Over 100 members originally expressed interest, and we currently have 83 surgeons participating at 52 IRB-approved sites. The surgeons are a near 50/50 mix of academic and private practice surgeons, adding to our generalizability of results [5].

In determining study design, we debated over a prospective cohort design versus a randomized trial. Ultimately, we believed we did not know a single critical variable to randomize and felt a cohort to determine predictors would best serve a revision series with its rich number of potential factors. As compared to primary reconstructions, additional variables involved all of the previous reconstruction issues such as tunnel position, widening, graft choice, fixation, meniscus and chondral procedures, etc. We thus chose a prospective longitudinal cohort similar to the Framingham study for cardiovascular disease many years ago [1].

Surgeon inclusion was based on AOSSM membership, attendance at an introduction meeting, and a willingness to follow the procedural issues identified in the SOP. This included utilizing a Musculoskeletal Transplant Foundation graft if allograft was chosen. In the introductory meetings, we determined that radiographs would be a critical data point and made a required and recommended x-ray list. Required baseline radiographs included a bilateral standing AP view and a full extension lateral view. Additional recommended views, if they were taken as part of their standard clinical care, included a bilateral bent knee weight-bearing view at 45° (Rosenberg), a bilateral patellofemoral view (Sunrise or Merchant), and a bilateral long leg standing x-ray for alignment.

After obtaining informed consent, the patient filled out a 13-page questionnaire that included questions regarding demographics, sports participation, injury mechanism, comorbidities, and knee injury history. Within this questionnaire, each participant also completed a series of validated general and knee-specific outcome instruments, including the Knee injury and Osteoarthritis Outcome Score (KOOS), the International Knee Documentation Committee Subjective form (IKDC), and the Marx activity rating scale. Contained within the KOOS was the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). This was filled out preoperatively and at planned postoperative follow-up points of 2, 6, and 10 years. Patients were paid \$20 for filling out the questionnaire each time. There were no other patient incentives since the treatment was standard care for a revision ACL reconstruction. Surgeons filled out a 42-page questionnaire at the time of the revision surgery that included the impression of the etiology of the previous failure, physical exam findings, surgical technique utilized, the intra-articular findings, and surgical management of meniscal and chondral damage.

Two-year patient follow-up was completed by mail with readministration of the same questionnaire as the one they completed at baseline. Patients were also contacted by phone to determine whether any subsequent surgery had occurred to either knee since their initial revision ACL reconstruction. If so, operative reports were obtained, whenever possible, in order to document pathology and treatment.

Completed data forms were mailed from each participating site to the data coordinating center. Data from both the patient and surgeon questionnaires were scanned with Teleform™ software (Cardiff Software, Inc., Vista, CA) utilizing optical character recognition, and the scanned data was verified and exported to a master database.

Teleform paper forms were utilized so data was available in real time. If starting today we would utilize electronic data capture, preferably REDCap, in 2006, it was felt this was best practice. Data was housed at Vanderbilt similar to MOON. Our data capture was relatively amazing

Table 42.1 Data completeness

Data completeness within a subset of variables (<i>n</i> = 900 cases)				
Outcome variable	Questionnaire source form	# of variables	# missing/observations (%)	% complete
Marx activity level	Patient	4	23/3600 (0.6%)	99.4
IKDC	Patient	19	38/17,100 (0.2%)	99.8
KOOS	Patient	42	259/37,800 (0.7%)	99.3
Current graft type	Surgeon	1	1/900 (0.1%)	99.9
Surgical technique	Surgeon	1	2/900 (0.2%)	99.8
Rehab factors	Surgeon	6	35/5400 (0.7%)	99.3

with >99% of all patient and surgeon data points completed when assessed on our first 900 patients (Table 42.1). We relied on MOON personnel originally, but quickly identified the need for a national coordinator to assist Laura Huston, and Amanda Braun joined our team and was centered at the coordinating center at Washington University in St. Louis. Our coordinator was initially funded by personal research funds and industry support that had been donated and through the AOSSM. To truly run this study with the personnel necessary depended upon a larger sustained grant process such as the NIH or Department of Defense, which was the impetus to pursue NIH funding. Once 2-year follow-up began, we added a full-time follow-up coordinator that contacts all patients to send out and obtain questionnaires and perform phone follow-up. Each site manages their site with their personal health care personnel or research personnel. Beyond initial enrollment there is little personnel involvement for the sites. Staff developed an electronic newsletter (Fig. 42.1) that arrived monthly to all surgeons and coordinators and acted as an impetus to stimulate patient enrollment listing total and monthly enrollment figures for surgeons. Everyone wanted to appear on the Top 10 list.

Keys that helped us early on were frequently based on our experience with MOON and included the use of conference calls and the ability to communicate by email which was relatively new at the time. The group's experience with IRB helped new sites get approval, and the knowledge regarding grants helped immensely in obtaining funding.

We developed a Scientific Advisory Board for advice. This eight-member Board meets at

least annually and provides advice and oversight for the study. This has included what research studies should be performed via an application form by members. This has been critical for governance issues. The makeup is balanced by geography, gender, and practice type (Table 42.2).

42.2 Findings of MARS

1215 patients were enrolled in the study. Median age was 26 with a range from 12 to 63 years. 505 (42%) were female. For 87% it was their first revision [5], while 13% were undergoing their second or higher revision ACL reconstruction. Seventy-three percent were injured while playing a sport. The most common sports were soccer and basketball that involve both genders. For the first decade of the modern ACL reconstruction with appropriate grafts and tunnel location, the etiology of ACL graft failure was felt to be technical issues [6] (Tables 42.3 and 42.4). In this cohort traumatic failure was felt to be more common, which may reflect two issues: (1) improved technical ability with improved training and education and (2) surgeons self-reporting on their own failures in this cohort. 334 (28%) were the surgeons own failures. Associated surgeries included high tibial osteotomies (*n* = 21), medial meniscus transplants (*n* = 34), and lateral meniscus transplants (*n* = 10).

The cause of technical failure was most commonly felt to be due to the femoral tunnel (Table 42.5) with the tibial tunnel the second most common cause. Seldom was fixation felt to be an issue. Varus and valgus malalignment was

MARS

MULTICENTER ACL REVISION STUDY

The Rover

Volume 4 (1) July 5, 2011

What's New?

- ◆ **Enrollment is now complete!** Please complete and send to Vanderbilt any MARS enrollment packets that have not yet been turned in. We would like to have **all completed packets turned in by 7/15/11.**
- ◆ **2 Year Follow Up Update Summary**
 - ⇒ 92% (23/ 25) Complete for year 2008
 - ⇒ 88% (125/ 142) Complete for year 2007
 - ⇒ 78% (227/ 292) Complete for year 2008
 - ◆ We need 22 more questionnaires for 85%.
 - ⇒ 27% (85/ 318) Complete for year 2009
 - ⇒ 75% (480/615) Complete Overall as of 7/1/11
- ◆ Thank you and farewell to Jennifer Sheehan, our 2 year follow up coordinator at Washington University. Jennifer has taken a new position outside of the University. Please email Amanda Haas regarding any follow up questions or concerns.
- ◆ Please call the individuals on your surgeon call sheets.

Upcoming Events

The 2011 AOSSM Annual Meeting will be held at the Manchester Grand Hyatt in San Diego, California, July 7-10, 2011.

The MARS meeting is scheduled to take place July 9th at 1:00 PM in the Mohsen AB room in the Manchester Grand Hyatt.

The MARS Scientific Advisory Board meeting will be held Saturday, July 9th at 7:00 AM in Lael's Restaurant at the Manchester Grand Hyatt.

Martians of the Month

Congratulations to Dr. Robert Marx and his coordinator, Brian Boyle, from The Hospital for Special Surgery (New York, NY) for enrolling the most patients this month (n=3).

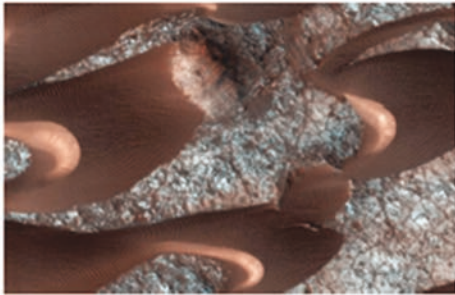
Reminders

- ⇒ Please submit the Facebook amendment. This information has been sent to the research coordinators.
- ⇒ We will send out updated call lists each month. If you don't receive a list, we have already received your participants' questionnaires.

Upcoming IRB Renewals

Lennox Hill—IRB Expires 7/9/11
 HSS—IRB Expires 7/14/11
 National Sports Med Institute—IRB Expires 7/18/11
 UNC—IRB Expires 7/18/11
 Arthro and Ortho Sports Med—IRB Expires 7/28/11
 Connecticut Children's—IRB Expires 8/3/11
 UCSF—IRB Expires 8/6/11
 3B Ortho/ U Penn—IRB Expires 8/12/11
 HealthPartners—IRB Expires 9/28/11
 Inland Ortho—IRB Expires 9/30/11
 Methodist Hospital—IRB Expires 10/12/11
 UHZ—IRB Expires 10/25/11
 Ortho Assoc of Aspen—IRB Expires 10/31/11
 University of Michigan—IRB Expires 11/3/11
 University of Virginia—IRB Expires 11/4/11
 Slocum Center—IRB Expires 11/4/11
 CHOP—IRB Expires 11/10/11
 Carrell Clinic—IRB Expires 12/15/11
 Uconn—IRB Expires 1/4/2012
 Mayo Clinic—IRB Expires 1/18/12
 University of Missouri—IRB Expires 1/22/12
 NYU—IRB Expires 1/23/12

Need New IRB Renewal Approvals From:
 Fowler-Kennedy
 COR
 Bridger Ortho
 Princeton Ortho
 University of Colorado



SANDS OF MARS
 Martian sand piles up in weird crescent-shaped dunes in the Nilil Fatera region of the Red Planet. These orbital pictures were captured by the high-resolution camera on NASA's Mars Reconnaissance Orbiter and released on June 9. (<http://www.msnbc.msn.com>)

Questions or Comments? E-mail us at haara@wudosis.wustl.edu or call (314) 362-3768

Fig. 42.1 Electronic Newsletter

Table 42.2 Time from last ACL reconstruction

	#	%
<1 year	149	12
1–2 years	389	32
>2 years	654	55
Unknown	8	<1
Total	1200	100

Table 42.3 Cause of failure

	#	%
Traumatic	671	56
Technical	615	51
Biologic	325	27
Other	35	3
Blank	2	<1

Note: 36% (427/1200) of these responses listed a combination of these components as reason for failure. The denominator is >100% due to the multiple choice option of this question (surgeons were instructed to “check all that apply”).

Table 42.4 Venn diagram of cause of failure

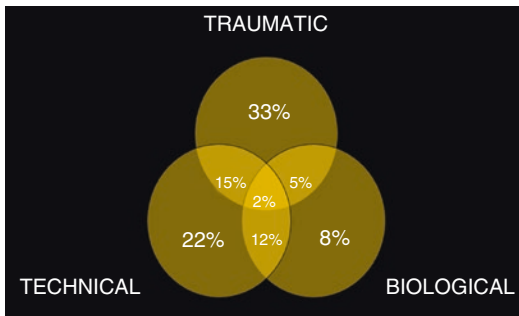


Table 42.5 Cause of technical failure

Femoral tunnel	595
Tibial tunnel	228
V/V malalignment	23
Femoral fixation	38
Tibial fixation	18
Autograft source	17
Allograft source	66
Posteromedial laxity	16
Posterolateral laxity	4

also seldom felt to be an issue. Prior approach was most commonly a single incision (80%) (Table 42.6). Two-incision or rear-entry approach was used in 17%. Prior graft utilized was

Table 42.6 Previous approach

	#	%
One incision	972	80
Two incisions	201	17
Trad'l arthrotomy	12	1
Mini arthrotomy	10	<1
Blank	5	<1
Total	1200	100

18 previous double bundle

Table 42.7 Previous graft source

	Autograft	Allograft
BTB	485 (40%)	133 (11%)
HS STG	285 (24%)	14 (1%)
HS ST	27 (2%)	3 (<1%)
Quad	4 (<1%)	
ITB	3 (<1%)	
Achilles		39 (3%)
Ant tib		61 (5%)
Post tib		13 (1%)
Other/unlk	5 (<1%)	73 (6%)
Blank		1 (<1%)
Combined	6 (<1%)	8 (<1%)
	815 (68%)	345 (29%)

Table 42.8 Current graft choice

	Autograft	Allograft
BTB	314 (26%)	285 (24%)
Quad	18 (2%)	3 (<1%)
HS ST	20 (2%)	17 (1%)
HS STG	219 (18%)	4 (<1%)
Achilles		83 (7%)
Ant tib		137 (11%)
Post tib		52 (4%)
Other/unlk	1 (<1%)	1 (<1%)
Blank		1 (<1%)
Combined	2 (<1%)	10 (<1%)
Total	574 (48%)	593 (49%)

Note: 3% used both auto- and allograft

autograft 68% of the time, allograft 29%, and a combination (autograft + allograft) 2%. The most common autograft was patellar tendon, utilized 40% of the time. Patellar tendon was the most common allograft utilized (11%) (Table 42.7). Current revision graft utilized was 26% patellar tendon autograft, 24% patellar tendon allograft, 22% soft tissue autograft, and 25% soft tissue allograft (Table 42.8).

42.2.1 Graft Choice

We analyzed graft choice as a predictor for outcome as one of the specific aims of our NIH-funded grant [4]. The demographics of patients that received autograft and allograft can be seen in Table 42.9. Allografts were placed in older, less active patients on average. The form of treatment was known for each allograft and included no radiation, light total body irradiation <1.8 mrad or rarely terminal radiation (Table 42.10). The overall re-rupture rate at 2 years was 37/1112 (3.3%), including 12 autograft, 24 allograft, and 1 combination graft. Autograft use was 2.78 times less likely to re-rupture ($p = 0.047$; 95% CI = 1.01, 7.69). No difference in re-rupture rate was found in soft tissue vs. BTB in either autograft or allograft. IKDC score improved with autograft reconstruction with an odds ratio of 1.33 (95% CI: 1.01–1.7;

$p = 0.045$). KOOS Sports and Recreation subscale significantly improved with autograft (OR 1.33; CI: 1.02–1.73; $p = 0.037$), as did the KOOS Quality of Life (OR 1.33; CI: 1.03–1.73; $p = 0.031$). The KOOS ADL and Symptoms subscales were not affected by graft choice.

MARS Findings

- Autograft 2.78 × less likely to re-rupture
- 91% of patients have meniscal or articular cartilage damage
- Previous lateral meniscectomy and trochlear groove chondrosis strong predictors of worse outcome
- Metal fixation predicted better outcome
- Early weight bearing, ROM did not impact outcome

Table 42.9 Graft choice demographics

	Autograft group (n = 584)	Allograft group (n = 601)	Auto + allo group (n = 34)
Males (%)	352 (60%)	337 (56%)	14 (41%)
Females (%)	232 (40%)	264 (44%)	20 (59%)
Median age (25%, 75% quartile)	24 (19, 32)	28 (21, 36)	22 (19, 31)
Median baseline Marx activity level (25%, 75% quartile)	12 (4, 16)	10 (3, 15)	11 (8, 16)
Median T2 Marx activity level (25%, 75% quartile)	8 (3, 12)	5 (1, 11)	10 (3, 15)

Many surgeons believed that graft choice was a predetermined fate in a revision setting and that the surgeon truly had no choice in determining what graft would be used for the patient. To analyze this belief, we performed a propensity analysis for graft choice (Fig. 42.2). Our analysis demonstrated that the surgeon performing the procedure was far and away the biggest factor on what graft type was chosen for the revision reconstruction, approximately five times more impactful than the second most common predictor (which was prior graft). Thus, surgeons truly did have a choice in what graft they utilized.

Table 42.10 Allograft treatment

Sterilization method	MARS allograft cohort	MARS allograft failure cohort
Aseptic	247 (42%)	13 (52%)
Whole body 1.2–1.8 mrad	313 (53%)	11 (44%)
Terminal 0.7–1.0 mrad	31 (5%)	1 (4%)

42.2.2 Meniscus and Articular Cartilage

Compared to primary reconstructions, the revision patient has a much higher chance of having meniscus or articular cartilage damage at the time of revision reconstruction. In our cohort, only 9% of the patients did not have a meniscus tear or grade 2 or worse articular cartilage damage. 91% had at least meniscus or articular cartilage damage, and 60% had both (Table 42.11). We analyzed the impact of these meniscus tears

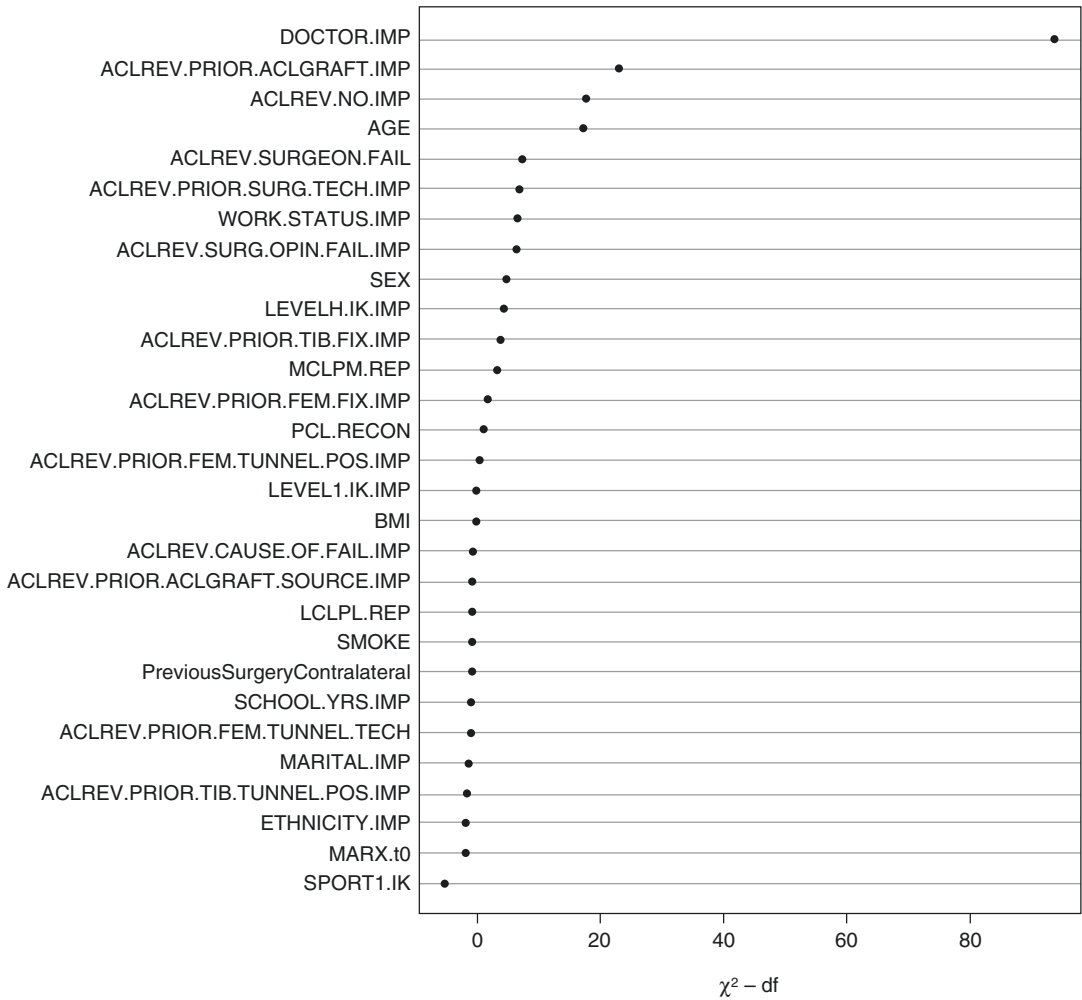


Fig. 42.2 Propensity values

Table 42.11 Meniscal and chondral damage

		Articular cartilage pathology	
		Normal	Abnormal
Meniscal pathology	Normal	109 (9%)	146 (12%)
	Abnormal	226 (19%)	719 (60%)

and articular cartilage damage on patient outcomes at 2 years [2]. Previous lateral meniscectomy prior to the time of revision significantly resulted in worse patient-reported outcomes (Table 42.12) and previous medial meniscectomy less so. Grade 2 or worse articular cartilage damage grade also impacted patient-reported outcomes at 2 years. The most significant finding

was the impact of trochlear groove chondrosis (Table 42.13).

42.2.3 Surgical Factors

Surgical factors were analyzed to determine impact on patient-reported outcome measures at 2 years [7]. A variety of factors were analyzed, and in many cases, it is difficult to determine intuitively why certain factors impacted outcome. With regard to surgical approach, having undergone a prior arthrotomy decreased IKDC scores ($p = 0.037$, OR = 2.43) and decreased all KOOS subscales ($p < 0.05$, OR range = 2.38–4.35).

Table 42.12 Meniscus impact on PROs

Structure	Marx	KOOS					IKDC	WOMAC		
		Symptoms	Pain	ADL	Sports-rec	QoL		Stiffness	Pain	ADL
Meniscus (previous pathology)										
Medial			0.002				0.035			
Lateral		0.008	0.042			<0.001	0.038	0.03	0.032	
Meniscus (current pathology)										
Medial										
Lateral										

Table 42.13 Articular cartilage impact on PROs

Structure	Marx	KOOS					IKDC	WOMAC		
		Symptoms	Pain	ADL	Sports-Rec	QoL		Stiffness	Pain	ADL
Articular cartilage (previous)										
Yes/no										
Articular cartilage (current)										
Medial femoral condyle			0.018						0.012	
Lateral femoral condyle				0.048						0.048
Medial tibial plateau					0.004					
Lateral tibial plateau										
Patella										
Trochlea			0.034	<0.001	0.011		0.01			<0.001

A double femoral tunnel resulted in worse KOOS QOL scores ($p = 0.027$, OR = 3.13). An ideal tibial position that was not enlarged resulted in worse KOOS, WOMAC, and IKDC scores ($p = 0.001$ – 0.03 , OR 1.19–2.68). Using a femoral tunnel declared “optimum” vs. drilling an entirely new femoral tunnel resulted in worse KOOS QOL scores ($p = 0.025$, OR 1.79). Undergoing a notchplasty decreased KOOS, IKDC, and WOMAC scores ($p = 0.013$ – 0.034 , OR 1.40–1.49). Factors that did not impact outcome included blended tunnels and knee position at time of graft fixation.

Graft fixation as a surgical factor impacting outcome was also analyzed. Femoral fixation with a metal screw had better 2-year outcomes in KOOS and WOMAC scores ($p = 0.01$ – 0.05 , OR 1.41–1.96) when compared to using a bioabsorbable screw, cross-pin, or combination. Tibial fixation other than metal screw resulted in worse IKDC ($p = 0.017$, OR 1.67) and WOMAC stiffness ($p = 0.013$, OR 1.72) scores.

Use of biologics and bone grafting was analyzed. Use of biologic enhancement in the revision setting resulted in lower 2-year MARX activity level scores ($p = 0.025$, OR 1.79). Utilizing femoral bone grafts resulted in lower MARX activity scores at 2 years ($p = 0.048$, OR = 2.04). Conversely, not bone grafting the tibia resulted in worse KOOS Pain scores ($p = 0.046$, OR 1.95) and WOMAC Pain scores ($p = 0.004$, OR 3.31).

42.2.4 Rehabilitation Factors

Rehabilitation factors were analyzed regarding their potential impact on revision ACL reconstruction outcomes [3]. Two rehabilitation factors predicted outcome: (1) use of an ACL derotation brace for return-to-sport had better KOOS sports/rec scores at 2 years (odds ratio = 1.50; 95% CI = 1.07–2.11; $p = 0.019$) and (2) use of an ACL derotation brace for postoperative rehabilitation period were 2.3 times more likely to have a

subsequent surgery by 2 years (OR = 2.26; 95% CI = 1.11–4.60; $p = 0.024$). Use of an ACL derotation brace at the time of return-to-sport could not be determined to improve or decrease the graft re-rupture rate. Restricting or allowing of all other factors did not predict outcome including active range of motion, passive range of motion, immediate weight-bearing, and the use of rehabilitative postoperative bracing.

42.3 Challenges

Obviously, a study of this magnitude involving this many centers and surgeons faces multiple significant challenges.

Challenges

- Maintaining IRB
- Funding
- Authorship
- Patient follow-up

42.3.1 Institutional Review Boards (IRB)

Obtaining and maintaining IRB approval at 52 sites take significant effort from the coordinators and sites. Each site required individual submissions, with their own forms and institutional requirements. No two forms or submission materials were identical. As such, this was quite time intensive in the beginning and continues to be a burden in ensuring that all sites maintain their active IRB status. A study of this magnitude initially needed at least one FTE to keep up with all 52 sites' submissions (and any subsequent modifications that were requested by each IRB) and was vastly underestimated by us. Recently the NIH has adopted that a single IRB be coordinated from a single site, and it is hopeful that this will decrease the IRB burden for large multicenter studies of this type. Some sites require a payment to obtain and maintain their IRB for this study, and this was a challenge from a funding standpoint initially.

42.3.2 Funding

Funding a study of this magnitude is probably our biggest challenge. Initially we relied upon start-up funds from industry which were funneled through AOSSM to support our coordinators, but this was stopped when one of our partners believed the study might cast their products in a poor light. An additional industry partner provided unrestricted funds that also supported our coordinator efforts. Ultimately, we knew we would need a large grant to support these efforts. Fortunately, we were able to obtain NIH funding in 2011 and have successfully extended that funding with a competitive renewal grant thru the middle of 2022. Department of Defense grants would have been another option or larger sustainable industry or society grants. Realistically, it takes those types of substantial grants to run these studies, but the findings can be more impactful than virtually any other type of efforts for the societies, and an investment in this type of study at start-up can pay off for the society with notoriety and ultimately a nationally funded study such as ours.

42.3.3 Variation of Surgeon Involvement

In hindsight one area that we would do differently with a future study would be to instill more rigorous rules from the start. I expected everyone to cooperate and work well together if they wanted to remain in the study and unfortunately did not establish enough ground rules for maintaining membership in MARS. It became apparent that there were surgeons that were not enrolling their patients after they received IRB approval from their institution. For example, they stated that they performed 25 revisions per year during the time of the surgeon training meetings but then had only enrolled 1 patient during their first 3 months after IRB approval. Because we had not established inclusion rules, it became difficult to police this type of activity or to prove that from an objective standpoint. In the future I would require all surgeon members to enroll at

least one patient and to require surgeon enrollment logs be submitted in order to verify that they were enrolling 80–90% of the patients in the study that were eligible. Additionally, I would ask and expect that they would assist in contacting and helping follow their patients.

42.3.4 Authorship

Authorship can be a cantankerous topic for members. Multiple issues exist for this topic. One issue is the assumption that when a few members take data that has been analyzed and by writing a manuscript reporting the findings that they should be the only recognized authors. This discounts the national coordinators and grant writers and Scientific Advisory Board members that have worked diligently behind the scenes to keep the study funded and up and running. Thus, a system where there would be four or five named authors and the MARS Group listed on the masthead became a problem. Thus, I adopted corporate authorship with everyone listed as an author that met author criteria, and I was able to convince the journals we typically utilize to make all acknowledged authors PubMed searchable. The order of the acknowledged authors is based on scientific contribution with first, second, and last making the most significant contribution in planning, writing, data analysis, etc. This has been effective in decreasing the concerns about why some were named on the masthead and others were not. It has to be acknowledged that this type of research and handling of authorship be recognized by department chairmen and promotion committees for academic members who need publications for academic progress. Chairmen need to realize this is rigorous Level 1 or 2 research and membership requires effort and participating in these studies is clinically meaningful and practice changing.

We have maintained strict standards on authorship. When a manuscript is complete, it is sent to all members with a deadline for returning the edited or approved manuscript. If the deadline is missed, then you are not listed on the final acknowledged author list for publication. This can be challenging for the authors who then need

to deal with multiple recommendations for improving the manuscript, but the final product is always improved.

42.3.5 Peer-Reviewed Journal Submissions

The reviews from orthopedic and sports medicine journals have been challenging. These are complex studies, and we have taken more than 80 potential reviewers out of commission with their MARS membership. It appears the concept of multivariable analysis is confusing. We try to explain that our variables are controlled in a way that they are independent predictors of outcome and not associated with other variables analyzed. For instance, if notchplasty is a predictor for worse outcome, then reviewers will comment it must be a surrogate for osteoarthritis or chondrosis in the rest of the knee. We explain that chondrosis was a variable and was controlled, but it comes up with every submission. We will continue to try to educate in our manuscripts.

42.3.6 Patient Follow-Up

Follow-up obviously is critical and the key to success of the study. It has been challenging to maintain this with each passing year. We currently have a full-time research assistant whose sole responsibility is reaching out to patients to keep up with their contact information and cajole them into filling out their follow-up questionnaires. We have found a simple mailing will achieve 40–50% return. Reaching out by the staff will get us to approximately 70% follow-up. To get into the 80% follow-up percentiles requires individual surgeons reaching out to patients, which is obviously time and work intensive, but has allowed us to stay above the 80th percentile for our follow-up. A challenge has been the sites that won't allow us to contact their patients for follow-up (due to the individual institution's IRB regulations or any non-US-based site), and we must rely on the sites to do the work. Their personnel are typically involved in several studies,

and the MARS patients may not be their top priority. Personally, it has been difficult enough that I would never use sites again in a multicenter study that required site vs. central follow-up.

42.4 Conclusions

While there have been multiple challenges encountered in the MARS study, they have for the most part been surmountable. The level of research and the questions that can be asked and answered in a cohort of this size and type is unmatched by any other approach. We believe the study design and scaffolding we have developed for this type of truly multi-surgeon multicenter research can be a model for future groups.

References

1. Cook NR, Moons KG, Harrell FE Jr. Assessing predictive performance beyond the Framingham risk score. *JAMA*. 2010;303:1368–9; author reply 1369.
2. Group M. Meniscal and articular cartilage predictors of clinical outcome following revision anterior cruciate ligament reconstruction. *Am J Sports Med*. 2016;44:1671–9.
3. MARS Group (Wright RW corresponding author). Surgical predictors of clinical outcome after revision anterior cruciate ligament reconstruction. *Am J Sports Med*. 2017;45(11):2586–94.
4. Group M, Group M. Effect of graft choice on the outcome of revision anterior cruciate ligament reconstruction in the Multicenter ACL Revision Study (MARS) Cohort. *Am J Sports Med*. 2014;42:2301–10.
5. Group M, Wright RW, Huston LJ, Spindler KP, Dunn WR, Haas AK, et al. Descriptive epidemiology of the Multicenter ACL Revision Study (MARS) cohort. *Am J Sports Med*. 2010;38:1979–86.
6. Johnson DL, Swenson TM, Irrgang JJ, Fu FH, Harner CD. Revision anterior cruciate ligament surgery: experience from Pittsburgh. *Clin Orthop Relat Res*. 1996;100–9.
7. MARS Group, Allen CR, Anderson AF, Cooper DE, DeBerardino TM, Dunn WR, et al. Surgical predictors of clinical outcomes after revision anterior cruciate ligament reconstruction. *Am J Sports Med*. 2017;45:2586–94.
8. Spindler KP, Kuhn JE, Freedman KB, Matthews CE, Dittus RS, Harrell FE Jr. Anterior cruciate ligament reconstruction autograft choice: bone-tendon-bone versus hamstring: does it really matter? A systematic review. *Am J Sports Med*. 2004;32:1986–95.
9. Spindler KP, Warren TA, Callison JC Jr, Secic M, Fleisch SB, Wright RW. Clinical outcome at a minimum of five years after reconstruction of the anterior cruciate ligament. *J Bone Joint Surg Am*. 2005;87:1673–9.
10. Wright R, Spindler K, Huston L, Amendola A, Andrish J, Brophy R, et al. Revision ACL reconstruction outcomes: MOON cohort. *J Knee Surg*. 2011;24:289–94.
11. Wright RW, Dunn WR, Amendola A, Andrish JT, Bergfeld J, Kaeding CC, et al. Risk of tearing the intact anterior cruciate ligament in the contralateral knee and rupturing the anterior cruciate ligament graft during the first 2 years after anterior cruciate ligament reconstruction: a prospective MOON cohort study. *Am J Sports Med*. 2007;35:1131–4.
12. Wright RW, Dunn WR, Amendola A, Andrish JT, Flanigan DC, Jones M, et al. Anterior cruciate ligament revision reconstruction: two-year results from the MOON cohort. *J Knee Surg*. 2007;20:308–11.
13. Wright RW, Gill CS, Chen L, Brophy RH, Matava MJ, Smith MV, et al. Outcome of revision anterior cruciate ligament reconstruction: a systematic review. *J Bone Joint Surg Am*. 2012;94:531–6.
14. Wright RW, Magnussen RA, Dunn WR, Spindler KP. Ipsilateral graft and contralateral ACL rupture at five years or more following ACL reconstruction: a systematic review. *J Bone Joint Surg Am*. 2011;93:1159–65.



Multicenter Study: How to Pull It Off? The PIVOT Trial

43

Eleonor Svantesson, Eric Hamrin Senorski,
Alicia Oostdyk, Yuichi Hoshino,
Kristian Samuelsson, and Volker Musahl

43.1 Introduction

One of the most important characteristics for a researcher is to question current practice and formulate hypotheses for further research contributing to evidence-based medicine. However, even the most interesting research question may be of no value to evidence-based medicine if inadequate methodology is chosen to study the questions proposed. The higher quality the methods used in the study, the more trustworthy will the results be.

Randomized controlled trials (RCTs) have for decades been considered as providing the highest level of evidence of all study designs. However, RCTs have also been criticized for not reflecting reality since they are often conducted only in highly specialized centers [17] and on a very selected study population after applying strict exclusion criteria. A multicenter design might increase the external validity of not only RCTs but also prospective and retrospective cohort trials. In fact, multicenter cohort trials are valuable complements to single-center RCTs, and the performance of such trials should be encouraged.

Multicenter trials offer many advantages. They enable investigation of large populations with different ethnicities and demographic characteristics and offer a possibility to compare results among participating centers, all important factors for increasing the generalizability of the study. However, to conduct a multicenter trial is challenging. The strengths of the study design are at risk of instead being limitations if the cooperation between the centers is not ensured and if the study protocol is not carefully prepared before study start, with minimal room for discrepancies of study performance across centers. This chapter focuses on sharing the experiences of how the Prospective International Validation of Outcome Technology (PIVOT) trial, a prospective multicenter trial, was planned and performed across four international centers and on presenting some of the outcomes and learning points of the trial.

E. Svantesson (✉)

Department of Orthopaedics, Institute of Clinical Sciences, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

E. H. Senorski

Department of Health and Rehabilitation, Institute of Neuroscience and Physiology, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

A. Oostdyk · V. Musahl

Department of Orthopaedic Surgery, University of Pittsburgh, Pittsburgh, PA, USA

Y. Hoshino

Department of Orthopaedic Surgery, Kobe University, Kobe, Japan

K. Samuelsson

Department of Orthopaedics, Institute of Clinical Sciences, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

Department of Orthopaedics, Sahlgrenska University Hospital, Mölndal, Sweden

Fact Box 43.1

Multicenter trials offer many advantages and have high external validity. The Prospective International Validation of Outcome Technology (PIVOT) trial was a prospective multicenter trial, planned and performed across four international centers.

43.2 The PIVOT Trial

The PIVOT trial was a multicenter study performed between four academic centers: the University of Pittsburgh (Pittsburgh, PA, USA), Istituto Ortopedico Rizzoli (Bologna, Italy), Sahlgrenska University Hospital (Gothenburg, Sweden), and Kobe University (Kobe, Japan).

The purpose of the study was to provide novel *in vivo* information about the validity of quantitative measurements of the pivot shift test to aid in the assessment of anterior cruciate ligament (ACL) deficiency and the outcome after ACL reconstruction. The goal of the trial was to provide a foundation for utilizing quantitative analysis of the pivot shift for ACL-deficient knees

before surgery and optimize ACL reconstruction for restoring dynamic knee function and ultimately improve patient outcome.

Two separate entities of the pivot shift test were quantified—tibial acceleration and lateral compartment tibial translation during the reduction phase of the pivot shift. For this purpose, two noninvasive technological devices were used. The tibial acceleration was quantified by an inertial sensor system (KiRA, Orthokey LLC, Lewes, DE, USA) consisting of a triaxial accelerometer that is held in place over Gerdy's tubercle, on the lateral aspect of tibia, by hypoallergenic skin straps (Fig. 43.1) [20, 21]. The sensor system communicates with a tablet PC via Bluetooth. The tibial translation was quantified via an image analysis system [7]. Three markers are placed on three specific landmarks at the lateral aspect of the knee, which are tracked via video recording using a commercial tablet (iPad, Apple Inc., Cupertino, CA) during the execution of the pivot shift test (Fig. 43.2). The specially designed software system subsequently analyzes the relative movement of the three markers and calculates the tibial translation relative to the femur and also displays this graphically [11].

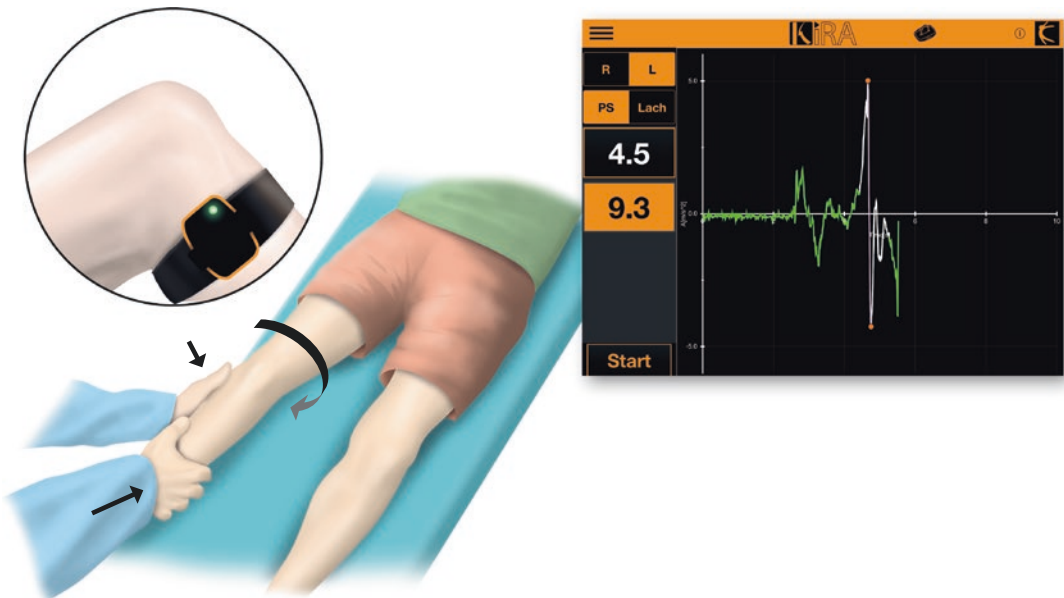


Fig. 43.1 The KiRA inertial sensor system for quantifying lateral tibial acceleration during the pivot shift test

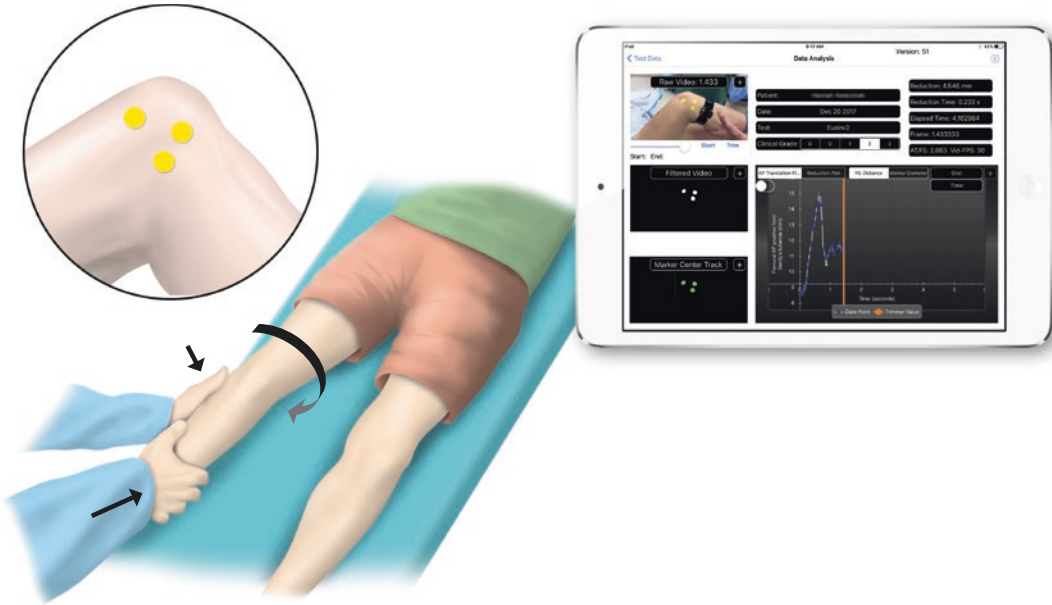


Fig. 43.2 Image analysis system on iPad for quantifying lateral tibial translation during the pivot shift test

Fact Box 43.2

The purpose of the study was to provide novel in vivo information about the validity of quantitative measurements of the pivot shift test to aid in the assessment of anterior cruciate ligament (ACL) deficiency and the outcome after ACL reconstruction. Two noninvasive technological devices were used, quantifying tibial acceleration and lateral compartment tibial translation during the pivot shift, respectively.

aspect of the future trial. In Bologna, the execution of the pivot shift was practiced, the two devices for quantification of the pivot shift test tested and ACL reconstruction was performed together to assimilate the procedure (Fig. 43.3). The upcoming sections highlight specific parts of study preparation and are important sections to consider when developing a manual of operations and procedures.

43.3 Preparation Prior to Study Start

The PIVOT trial was preceded by a phase of detailed study preparation, including a series of important steps to ensure performance of a high-quality trial. A corner stone in the early study preparation was the investigators meeting held in Bologna in 2012 with investigators from all institutions present, where the Manual of Operations and Procedures was unanimously developed, which was fundamental for every

43.4 Establishing Aims and Specific Research Questions

When setting up a study, it is essential that the aim and hypothesis are clearly formulated together with the specific research questions prior to start of the study. These recommended points are all fundamental for how the process of study planning continues since the next steps include choosing an appropriate methodology to analyze the results and answer the predefined research questions. Unfortunately, studies that are not based on these fundamentals are at risk of being subjected to “fishing” for interesting results rather than rely-

Fig. 43.3 The PIVOT team practicing the pivot shift test in Bologna, 2012



ing on testing the hypothesis. As honest researchers, we need to be aware that this type of research is biased, and also, that this type of approach to research surely exists among published literature. In a multicenter trial, it is important to clearly frame the purpose of the study. If all participating centers are in agreement regarding the aims of the study and what is expected and relevant prior to study start, not only can unnecessary disputes and misunderstandings be avoided, but it also ascertains that contribution to publication bias is limited. It is important that the purpose and research questions are not too vague, but clear and answerable. The agreement of the research questions in the PIVOT trial was accomplished through an open communication in which all centers participated and shared their opinions until a consensus was reached. Thereafter, the aims and research questions were documented in the manual of operations and procedures to function as the overall common guide during study performance. The PIVOT trial had three specific aims, all accompanied with a hypothesis and specific research questions.

Another important factor for a well-functioning collaboration was that it was decided what aim each center was primarily responsible for, i.e., each center was appointed to take the lead in the manuscript writing for specific aims from the

beginning. To be clear about this before knowing the results was, from our point of view, important for promoting a collaborative environment and avoid potential future friction in the group.

Fact Box 43.3

It is essential that the aim and hypothesis are clearly formulated together with the specific research questions prior to start of the study. The agreement of the research questions in the PIVOT trial was accomplished through an open communication in which all centers participated and shared their opinions until a consensus was reached.

43.5 Recruitment Plan and Study Population

As for all studies, the study protocol needs to be clear regarding the inclusion and exclusion criteria. Furthermore, a standardized way of recruiting patients needs to be established in order to avoid selection bias. It is important to acknowledge that each center may have different routines for this purpose, especially in a situation where centers from three continents are included.

Influencing factors may include legal policies, cultural factors, volume of the patients, and the participating researchers' own preferences. Therefore, this part of the study was discussed thoroughly among the participating sites with focus on how to include a population that could answer our research questions while eliminating as many confounding factors as possible. Strict inclusion and exclusion criteria were written down in the manual of operation. Moreover, a standardized screening process was established and a power analysis, taking into account an anticipated loss to follow-up, was performed to yield the total number of patients to be recruited. In agreement of the collaborators, a definition of evaluable patients was established, and it was also decided on how to handle patient withdrawal. As clinicians we are all aware of the hectic clinical work and the limited time for patient consultation, which could contribute to accidental deviation from the strict eligibility criteria. To minimize such events, and to further standardize the recruitment screening, special interview forms were developed as well as a checklist for inclusion and exclusion criteria.

Fact Box 43.4

A standardized way of recruiting patients needs to be established in order to avoid selection bias. The manual of operation included strict inclusion and exclusion criteria, a definition of evaluable patients, and how to handle patient withdrawal. To further standardize the recruitment screening, special interview forms were developed as well as checklists.

43.6 Time Line and Follow-Up Schedule

It was decided that the PIVOT trial would examine patients over a 24-month period. An important factor for the study purpose was that specific and clearly stated endpoints were formulated for each aim of the study. Without defined study endpoints, it is simply not possible to conduct a

good study since potential bias may be introduced. It was decided that all patients would be evaluated at 3, 6, 12, and 24 months after ACL reconstruction, and a detailed section for scheduled follow-up evaluation was written. For each follow-up occasion, an exact list of what variables to collect and how to collect these was created. All sites obtained standardized questionnaires and protocols for collection of data. Moreover, specific information of all patient-reported outcome measures and clinical tests were distributed to the clinics. Due to the language differences among each country in the PIVOT trial, forms to be completed by the patients were translated. A specific person was assigned at the lead site to periodically check the status of follow-up in each center and send notices for upcoming follow-up patients. This person plays a key role to ensure that no patients are missed during the follow-up phase of the study.

Fact Box 43.5

Specific and clearly stated endpoints were formulated for each aim of the study. It was decided that all patients would be evaluated at 3, 6, 12, and 24 months after ACL reconstruction, and a detailed section for scheduled follow-up evaluation was written.

43.7 Standardization of Clinical Testing

A multicenter trial entails that several examiners are involved in the trial. Therefore, it is important to undertake every possible action to standardize the exams and to have a common approach to the clinical testing. When planning for the PIVOT trial, the main goal was to create a standardized approach for data collection. A lot of effort has been undertaken to ensure that a strict training plan existed for clinicians that were to be involved in the examination of patients. The most challenging clinical testing facing the PIVOT trial was the execution of the pivot shift test.

The pivot shift test is a dynamic knee laxity test which has been referred to as the most specific test for detection of ACL deficiency [2]. However, there are several execution techniques described for the test [1, 14, 16], which entails that the test is limited by intra- and inter-variability [8, 9]. In an attempt to overcome this, we described and analyzed the results of a standardized execution technique for the pivot shift test among surgeons prior to study start [6, 14]. Moreover, a separate section in the manual of operations and procedures was written to describe the training plan, clearly stating the training requirements for participation. When the PIVOT trial subsequently was started, each examiner was provided a video outlining the steps of the standardized pivot shift test as well as written instructions for the performance. Additionally, the investigators involved in the PIVOT study underwent training for the pivot shift test during the University of Pittsburgh Panther Global Summit on ACL Reconstruction held in Pittsburgh in August 2011. Apart from the specific training and standardization of the pivot shift test, all examiners received written instructions for how to perform all other examinations in a standardized manner. Finally, all investigators and team members of the PIVOT trial had to complete Good Clinical Practices Training prior to the start of the study.

Fact Box 43.6

A lot of effort was undertaken to ensure that clinical examination was standardized among clinicians that were to be involved in the examination of patients. All investigators received oral and written instructions of the clinical exams and also underwent training prior to study start.

institution has an internal process for reviewing research trials for ethical purposes. With a study across four international centers, the legal standards regarding research conducted in each country must be addressed as well. Additionally, with studies recruiting human subjects, issues of informed consent must be addressed. For the PIVOT trial, the lead site developed protocols, data collection forms, and an informed consent template to be used to obtain all necessary approvals at each site.

It cannot be underlined enough that ensuring a system for documentation prior to start of the study is absolutely crucial for a multicenter study. Not only is this a matter of legal and ethical responsibility, but it also involves the database of your collected data and the documentation of communication between centers. One central database was used for all data collection to ensure standardization of all forms completed for the study. Make sure that there are persons appointed to take responsibility and coordinate documentation, and also ascertain that everyone involved in entering data are educated in how the documentation system works with training and written instructions. Furthermore, arrangements for how to document correspondence, teleconferences, and videoconferences among the sites need to be made. If everything is documented, the room for con-
tempers is minimized.

Fact Box 43.7

The legal standards regarding research conducted in each country were addressed, as well as issues of informed consent. One central database was used for all data collection to ensure standardization of all forms completed for the study. Furthermore, arrangement for how to document correspondence, teleconferences, and videoconferences among the sites was made.

43.8 Regulatory Responsibility and Trial Documentation

To conduct a clinical study, ethical concerns and the regulatory process must be addressed. Each

43.9 Communication Plan

Due to the nature of an international multicenter research study, it is imperative for investigators to have an open and consistent communication. The investigators of the PIVOT trial discussed and reached consensus regarding how communication should primarily take form and established a policy around this. All investigators agreed on scheduling teleconferences or videoconferences at least monthly, as well as more frequent communication via email to address questions and issues as they emerged.

Investigator meetings were also a prioritized matter among the investigators. During the planning phase of the PIVOT trial, the study investigators met to discuss and finalize plans for implementation of the study. The meeting included review and approval of the manual of

operations and procedures, including a final confirmation of what had been decided in terms of; effective and culturally appropriate recruitment and retention strategies, standardized surgical and outcome measure procedures, plans for training, approval of study forms, discussion of translating study forms, regulatory requirements, and any other necessary administrative activities that were needed for the successful and timely startup of the PIVOT trial. It was also decided on beforehand that the investigators would meet in person at least annually which could, for example, be in conjunction to an international meeting (Fig. 43.4). All in-person meetings were documented and minutes distributed for all members of the group that were unable to attend. Moreover, over the course of four years, the research team got together once at each participating center (Fig. 43.5).



Fig. 43.4 PIVOT team meeting at the 15th ESSKA Congress in Geneva, 2012



Fig. 43.5 Annual PIVOT meeting in Kobe, Japan

Fact Box 43.8

The investigators of the PIVOT trial discussed and reached consensus regarding how communication should primarily take form and established a policy around this. All in-person meetings were documented and minutes distributed for all members of the group that were unable to attend.

Fact Box 43.9: Key Points for a Successful Performance of a Multicenter Trial in the Experience of the PIVOT Trial

- Frame the purpose of the study and establish research questions and hypotheses prior to study start.
- Define how specific responsibilities should be distributed across the partici-

pating sites and prepare a plan for manuscript writing early in the planning process.

- Apply a standardized methodology across all parts of study performance including recruitment of patients, the screening process, the intervention, and the clinical examination. The use of checklists and protocols promotes a standardized approach.
- Document everything! Have persons appointed to take responsibility and coordinate documentation.
- Have an open and consistent communication. A multicenter trial is a teamwork, and everyone must contribute to a collaborative environment.
- Follow all regulatory requirements of each institution and country, respectfully.

43.10 The Results of the PIVOT Trial

43.10.1 Data Collection and Analysis

The collection of data was evaluated continuously during the enrollment phase. The enrollment phase was ended when having enrolled a total of 107 patients, which subsequently were followed for 24 months. As part of the study protocol, the statistical analysis for each study aim was planned in advance, and the setup of a common dataset facilitated data analysis. The principal site (University of Pittsburgh) took the lead in performing data analysis to increase consistency, ensuring that the investigators who performed the statistical analysis were well-familiar with the database. The PIVOT trial team followed the initial plan of which center that was responsible for writing the manuscript of each specific study aim. Additionally, a progress report was updated successively to keep track of study progress, publications, and assignments. All adverse events, reoperations, and contralateral ACL ruptures during the study period were also documented.

Fact Box 43.10

The statistical analysis for each study aim was planned in advance, and the setup of a common dataset facilitated data analysis. A progress report was updated successively to keep track of study progress, publications, and assignments.

43.10.2 Presentation of the Results

The first results presented from the PIVOT trial was a validation study of the noninvasive technology for quantitative pivot shift [13]. The paper concluded that both techniques were valid and that the technology was able to detect differences between clinically graded low- and high-grade pivot shift, findings which were indeed encourag-

ing for continued analysis of the data. Since then, a total of four other papers have been published in various journals [3, 10, 18, 19], and additional manuscripts are either submitted or in preparation. Over 20 presentations of the PIVOT trial have been held at international meetings, and several abstracts and posters have been presented. The PIVOT trial has also received attention as newsletter articles [4, 5] and as a podcast episode by the American Journal of Sports Medicine [12]. In 2017, the book *Rotatory Knee Instability: An Evidence Based Approach* edited by the principal investigator of each participating center, was published [15]. The book could be seen as a deepened overview of all major aspects of the assessment of rotatory knee instability, including the pivot shift phenomenon, and highlights current knowledge as well as future aspects that need to be addressed by further research.

Fact Box 43.11

A validation study of the noninvasive technology for quantitative pivot shift was primarily performed. Since then, a total of four other papers have been published and several manuscripts are in progress. Over 20 presentations of the PIVOT trial have been held at international meetings, and in 2017, the book *Rotatory Knee Instability: An Evidence Based Approach* was published.

Fact Box 43.12: Key Findings from the PIVOT Trial

- Both devices for quantification of the pivot shift were found valid and able to detect differences between clinically graded low- and high-grade pivot shift.
- Quantitative pivot shift detected a significantly higher tibial acceleration and lateral compartment translation in patients under anesthesia compared with in awake state.

- Preoperative tibial acceleration and lateral compartment translation were significantly reduced by anatomic single-bundle ACL reconstruction.
- Generalized joint laxity does not appear to correlate with quantitative pivot shift in the ACL-injured knee.
- Static anteroposterior knee laxity tests are poorly correlated to quantitative pivot shift in the ACL-injured knee. Static and rotatory knee joint laxity should be considered as separate entities of the knee examination.

43.11 Future Directions

During the course of the PIVOT trial, the collaboration among the centers has grown stronger, and the network has expanded. Conducting a multicenter trial has encouraged us to continue applying this methodology, i.e., quantitative evaluation of the pivot shift test, for high-quality research. Some of the preliminary research questions have been answered; however, they have also resulted in identification of new areas of research that need to be undertaken. For example, the results from the PIVOT trial have led to the setup of another study which aims to compare the outcomes and the rotatory laxity between ACL reconstruction and ACL reconstruction complemented by tenodesis.

Take-Home Message

- The multicenter PIVOT trial has provided novel knowledge about the use of quantitative pivot shift in ACL-injured patients across three continents.
- Performance of a multicenter trial requires teamwork, which is promoted by an open and consistent communication.
- A cornerstone for the successful execution of the PIVOT trial was the stringent preparation prior to study start, where consensus was reached across all participating sites for a standardized methodology regarding recruitment of patients, the screening process, the intervention, and the clinical examination.

References

1. Anderson AF, Rennert GW, Standeffer WC Jr. Clinical analysis of the pivot shift tests: description of the pivot drawer test. *Am J Knee Surg.* 2000;13(1):19–23.
2. Benjaminse A, Gokeler A, van der Schans CP. Clinical diagnosis of an anterior cruciate ligament rupture: a meta-analysis. *J Orthop Sports Phys Ther.* 2006;36(5):267–88. <https://doi.org/10.2519/jospt.2006.2011>.
3. Hamrin Senorski E, Svantesson E, Sundemo D, Musahl V, Zaffagnini S, Kuroda R, et al. Preoperative knee laxity measurements predict the achievement of a patient-acceptable symptom state after ACL reconstruction: a prospective multicenter study. *J ISAKOS: Joint Disord Orthop Sports Med.* 2018;3:26–32.
4. Hoshino Y, Musahl V, Kuroda R, Zaffagnini S, Samuelsson K, Lopomo N, et al. PIVOT Study Group Project Update ISAKOS Newsletter. 2014.
5. Hoshino Y, Musahl V, Ryosuke K, Zaffagnini S, Samuelsson K, Lopomo N, et al. Quantified pivot shift test by accelerometer and iPad App. ESSKA Newsletter. 2015.
6. Hoshino Y, Araujo P, Ahlden M, Moore CG, Kuroda R, Zaffagnini S, et al. Standardized pivot shift test improves measurement accuracy. *Knee Surg Sports Traumatol Arthrosc.* 2012;20(4):732–6. <https://doi.org/10.1007/s00167-011-1850-0>.
7. Hoshino Y, Araujo P, Ahlden M, Samuelsson K, Muller B, Hofbauer M, et al. Quantitative evaluation of the pivot shift by image analysis using the iPad. *Knee Surg Sports Traumatol Arthrosc.* 2013;21(4):975–80. <https://doi.org/10.1007/s00167-013-2396-0>.
8. Kim SJ, Kim HK. Reliability of the anterior drawer test, the pivot shift test, and the Lachman test. *Clin Orthop Relat Res.* 1995;(317):237–42.
9. Kuroda R, Hoshino Y, Kubo S, Araki D, Oka S, Nagamune K, et al. Similarities and differences of diagnostic manual tests for anterior cruciate ligament insufficiency: a global survey and kinematics assessment. *Am J Sports Med.* 2012;40(1):91–9. <https://doi.org/10.1177/0363546511423634>.
10. Lopomo N, Signorelli C, Rahnama-Azar AA, Raggi F, Hoshino Y, Samuelsson K, et al. Analysis of the influence of anaesthesia on the clinical and quantitative assessment of the pivot shift: a multicenter international study. *Knee Surg Sports Traumatol Arthrosc.* 2017;25(10):3004–11. <https://doi.org/10.1007/s00167-016-4130-1>.
11. Muller B, Hofbauer M, Rahnama-Azar AA, Wolf M, Araki D, Hoshino Y, et al. Development of computer tablet software for clinical quantification of lateral knee compartment translation during the pivot shift test. *Comput Methods Biomech Biomed Engin.* 2016;19(2):217–28. <https://doi.org/10.1080/10255842.2015.1006210>.
12. Musahl V. Validation of quantitative measures of rotatory knee laxity. *Am J Sports Med.* 2016;44(9):2393–8.

13. Musahl V, Griffith C, Irrgang JJ, Hoshino Y, Kuroda R, Lopomo N, et al. Validation of quantitative measures of rotatory knee laxity. *Am J Sports Med.* 2016;44(9):2393–8. <https://doi.org/10.1177/0363546516650667>.
14. Musahl V, Hoshino Y, Ahlden M, Araujo P, Irrgang JJ, Zaffagnini S, et al. The pivot shift: a global user guide. *Knee Surg Sports Traumatol Arthrosc.* 2012;20(4):724–31. <https://doi.org/10.1007/s00167-011-1859-4>.
15. Musahl V, Jón K, Kuroda R, Zaffagnini S. *Rotatory knee instability: an evidence based approach*: Springer International Publishing; 2017.
16. Noyes FR, Grood ES, Cummings JF, Wroble RR. An analysis of the pivot shift phenomenon. The knee motions and subluxations induced by different examiners. *Am J Sports Med.* 1991;19(2):148–55. <https://doi.org/10.1177/036354659101900210>.
17. Rosenberg W, Donald A. Evidence based medicine: an approach to clinical problem-solving. *BMJ.* 1995;310(6987):1122–6.
18. Sundemo D, Blom A, Hoshino Y, Kuroda R, Lopomo NF, Zaffagnini S, et al. Correlation between quantitative pivot shift and generalized joint laxity: a prospective multicenter study of ACL ruptures. *Knee Surg Sports Traumatol Arthrosc.* 2018;26(8):2362–70. <https://doi.org/10.1007/s00167-017-4785-2>.
19. Svantesson E, Hamrin Senorski E, Mårtensson J, Zaffagnini S, Kuroda R, Musahl V, et al. Static anteroposterior knee laxity tests are poorly correlated to quantitative pivot shift in the ACL-deficient knee: a prospective multicentre study. *J ISAKOS: Joint Disord Orthop Sports Med.* 2018;3:83–88.
20. Zaffagnini S, Lopomo N, Signorelli C, Marcheggiani Muccioli GM, Bonanzinga T, Grassi A, et al. Innovative technology for knee laxity evaluation: clinical applicability and reliability of inertial sensors for quantitative analysis of the pivot-shift test. *Clin Sports Med.* 2013;32(1):61–70. <https://doi.org/10.1016/j.csm.2012.08.007>.
21. Zaffagnini S, Signorelli C, Grassi A, Yue H, Raggi F, Urrizola F, et al. Assessment of the pivot shift using inertial sensors. *Curr Rev Musculoskelet Med.* 2016;9(2):160–3. <https://doi.org/10.1007/s12178-016-9333-z>.



Conducting a Multicenter Trial: Learning from the JUPITER (Justifying Patellar Instability Treatment by Early Results) Experience

Jason L. Koh, Shital Parikh, Beth Shubin Stein,
and The JUPITER Group

44.1 Introduction

Multicenter trials are critically important in answering significant research questions in orthopedic surgery. Due to the nature of orthopedic injuries and treatment, it can be difficult for a single center to accumulate enough patients to

JUPITER Group—Elizabeth A. Arendt, MD; Jackie Brady, MD; Dennis Crawford, MD; Diane L. Dahm, MD; Henry Ellis, MD; Matthew Halsey, MD; Peter Fabricant, MD; Jack Farr, MD; Dan Green, MD; Benton Heyworth, MD; Kosmas Kayes, MD; Dennis Kramer, MD; Aaron Krych, MD; Robert Magnussen, MD; Todd Milbrandt, MD; Matthew Milewski, MD; Charlie Popkin, MD; Lauren Redler, MD; David Roberts, MD; Verena Schreiber, MD; Seth Sherman, MD; Sabrina Strickland, MD; Marc Tompkins, MD; Eric Wall, MD; Philip Wilson, MD; Yi-Meng Yen, MD

The JUPITER Group

J. L. Koh (✉)

Department of Orthopaedic Surgery, NorthShore University Health System, Northshore Orthopedic Institute, Evanston, IL, USA

Pritzker School of Medicine, University of Chicago, Chicago, IL, USA

S. Parikh

Sports Medicine, Division of Pediatric Orthopaedic Surgery, Cincinnati Children's Hospital, University of Cincinnati, Cincinnati, OH, USA

B. S. Stein

Hospital for Special Surgery, Weill Cornell Medical College, New York, NY, USA

have sufficient statistical power to address a clinical question. In addition, the patient population or the particular practice patterns of a single location can make generalizing the results of a single-site study difficult [3]. For example, in patellofemoral instability, surgical results based on an injury group consisting primarily of traumatic injuries to male military recruits in their 20s may not be applicable to atraumatic dislocations in skeletally immature female patients with trochlear dysplasia. A multicenter trial might help address diverse patient populations and practice patterns.

Multicenter trials can provide important information by collecting sufficient numbers and varieties of patients to allow significant statistical power and generalizability. Challenges are related to the geographic separation and different locations that make trial coordination more difficult than in a single-site study.

There are unique challenges related to multicenter trials. The obvious one is that investigators are geographically spread apart, resulting in increased difficulty in communication and potential for increased variability in conducting the

study. What is just as critical is the dedication of the group of investigators to be willing to be collaborative and compromise to achieve a collective goal.

44.2 Initiation of Multicenter Trial

Typically, a multicenter trial begins with a small group of investigators who share a common interest in a clinical research question. This is often based on results of a single institution's experience and the desire to identify how this may be further generalizable. In the JUPITER trial, the investigators have been motivated to initiate a multicenter study investigating the results of treatment of patellofemoral instability in a pediatric, adolescent, and young adult population. The standard of care for initial acute patellofemoral dislocation has historically been nonoperative [10]; however, it has been demonstrated that the rate of recurrent instability can be quite high and also that many patients have symptoms or loss of function related to the dislocation. In addition, recurrent instability has been associated with a significant rate of articular cartilage damage and long-term osteoarthritis [6]. Recent work done at several centers has identified specific risk factors for recurrent instability [1, 2, 5, 9]. Algorithms have been proposed for treatment [11], but questions still remain as to the natural history and results of treatment for different patients [7, 8]. JUPITER is a multicenter, multi-armed prospective cohort study aimed at addressing some of these questions, particularly which patients can do well with an isolated medial patellofemoral ligament reconstruction for stabilization and which patients need other procedures.

Initiation of the JUPITER trial was based on a pilot single-center study and was designed to identify risk factors for recurrent patella instability and treatment outcomes.

44.3 Discussion and Planning Phase

A discussion phase was initiated after some initial face-to-face and email contacts among members of a group of researchers interested in the area of patellofemoral instability, primarily as determined by attendance at the International Patellofemoral Study Group (IPSG) meeting in Chicago in 2015. Interest was gauged by a small group of initial investigators. In most multicenter trials, there is a project leader or leadership team that helps keep the trial on track. Key to the conduct of this and any multicenter study is significant commitment by this group of investigators.

Discussion and planning began with a small group. Essential statistical evaluation was performed, and a screening form was developed and circulated to evaluate site and investigator capabilities.

The research goals were identified, and after this was clarified, a statistician performed a power calculation for the study. A statistician or epidemiologist with statistical training is a critical partner in identifying the appropriate number of patients to ensure sufficient power to answer the proposed research question. Appropriate corrections should be made for patient dropout or loss to follow-up. The calculation of overall enrollment numbers will be compared to anticipated individual site enrollments for a given period of time to help determine the total number of sites and/or anticipated length of time for enrollment.

A screening form for potential sites interested in joining the study was developed and circulated. In this tool, sites and investigators provided information about level of interest, site and investigator experience and support (including financial support and research personnel such as research assistants/coordinators), estimated frequency of enrollment, and anticipated level of commitment. A sample screening form page from JUPITER is shown in Fig. 44.1 (Tables 44.1 and 44.2).

JUPITER (Justifying Pediatric Instability Treatment by Early Results)
 SCREENING FORM

Name of surgeon: _____

Institution Name: _____

Affiliated University: _____

Phone No _____

Research Coordinator _____

Years in Practice: _____ years

Type of Practice: _____

Average no of Patellar Instability treated non-operatively per year: _____

Average no of Patellar Instability treated operatively per year: _____

Average no of Medial-sided repair per year: _____

Average no of isolated MPFL reconstructions per year: _____

Average no of TTO (Elmslie-Trillat, AMZ, distalization) per year: _____

Average no of Osteochondral fracture Rx following patellar stabilization per year: _____

Average no alignment osteotomies (femur/tibia, coronal/rotational) per year: _____

1. Of all operative patellar stabilization, how often do you do knee arthroscopy?

Knee Arthroscopy _____%

2. Of all operative patellar stabilization, how often patients have open femoral physis?

Open Physis _____%

Fig. 44.1 JUPITER Screening Form

44.3.1 Protocol Development

Protocol development was initially performed by the executive committee based on a pilot study initiated at one site. Multiple questions need to be answered during protocol development, including eligibility criteria and assessment. Assessment for clinical projects can include history, physical examination, and radiographic studies, as well as

patient-reported outcome scores and standardized evaluation tools.

The specific aims of JUPITER were to evaluate the safety and effectiveness of (1) nonoperative treatment, (2) isolated medial patellofemoral ligament (MPFL) reconstruction, and (3) MPFL reconstruction combined with bony procedures (osteotomy, trochleoplasty). Subject recruitment was planned for a 1 year time period at ten

Table 44.1 JUPITER authorship criteria (adapted from PRISM <https://www.prismsports.org/>)

Eligibility criteria for authorship in JUPITER manuscripts (adapted from PRISM)	
I. All of the following criteria must be met to be considered for authorship:	
1.	Maintain good standing in JUPITER, as defined in the manual of operations
2.	Respond with all of the following within 2 weeks for each manuscript
	(a) Comments/edits of manuscript (or an “all good” response)
	(b) Completion of all disclosure forms
	(c) Completion of all copyright transfer forms, etc.
II. In addition, investigators must have met a set of the following criteria (by receiving at least 3 points)	
1.	Participated in protocol development and study design—1 point
2.	Participated in writing the original manuscript—2 points
3.	Reviewed a rough draft of the article with substantial suggestions and editing—1 point
4.	Patient enrollment with complete data used for this study:
	(a) 1–9% of patients in study with sufficient follow up data—1 point
	(b) 10–29% of patients in study with sufficient follow up data—2 points
	(c) ≥30% of patients in study with sufficient follow up data—3 points
5.	Participated in grant writing for study group funding—2 points
III. Authorship order will be determined by Executive Committee based upon:	
	• Good standing in JUPITER
	• Amount and quality of manuscript drafting/edits/review
	• Number of patients entered into the registry
IV. If you wish to perform a research sub-study utilizing the multicenter database, you need to complete a “Research Proposal Form”	
	• The form will be reviewed by the investigators and coordinators at Cincinnati Children’s Hospital and the Hospital for Special Surgery to assure:
	– No conflict with existing study proposals
	– Compliance with the “FINER” criteria: Feasible, Interesting, Novel, Ethical, and Relevant
	• If the study is approved, all participating Investigators in the group will be notified about the study. Centers with clean and complete data related to the topic will be invited to participate
	• Preliminary authorship criteria and order will be established with a PI and one other representative from the proposing institution and a PI only from other participating institutions. The tag line “JUPITER study group” will be added to the list of authors on all publications

JUPITER-Version 1—February 2018

centers. Posttreatment outcome assessment was to be performed at 6, 12, and 24 months, including assessment of function, activity level, health-related quality of life, patellar stability, knee motion, and complications.

44.3.2 Clinical Assessment

In JUPITER, a draft assessment tool for data collection was developed. The initial tool was relatively lengthy, and multiple conference calls were made by the group of investigators to help further refine and develop the protocol. Critically, it was felt that it was important to simplify the initial form to minimize the burden on the investigators located at multiple sites. A simplified assessment tool also allows for improved patient compliance

and reproducibility of data. Validated tools are important to use to make sure the data appropriately reflects desired outcome evaluation; we use Pedi-IKDC, Kujala, HSS Pedi-FABS, Banff Patellofemoral Instability instrument 2.0, and KOOS Knee survey. Initially, the assessment tool was a paper document; however, specialty society grant funding was received allowing investigators to use a Web-based system for data collection and management (Oberd™, Columbia, Missouri USA), and the study transitioned to this during the course of enrollment. Other investigators have used REDCap (Research Electronic Data Capture), which is a free, research data management system sponsored by Vanderbilt University and supported by the National Institutes of Health. The advantages of using electronic databases are multiple: (1) can allow for remote col-

Table 44.2 JUPITER institutions and investigators

Institutions	Investigators
Cincinnati Children’s (Coordinating Center)	Shital Parikh, PI Eric Wall
Hospital for Special Surgery (Coordinating Center)	Beth Shubin Stein Dan Green Sabrina Strickland Peter Fabricant
Boston Children’s	Yi-Meng Yen Dennis Kramer Benton Heyworth Matthew Milewski
Columbia University	Charlie Popkin Lauren Redler
Mayo Clinic	Diane L. Dahm Todd Milbrandt Aaron Krych
NorthShore University Health System	Jason L Koh David Roberts Verena Schreiber
OrthoIndy	Jack Farr Kosmas Kayes
Oregon Health Sciences University	Jackie Brady Dennis Crawford Matthew Halsey
Ohio State University	Robert Magnussen
Texas Scottish Rite	Henry Ellis Philip Wilson
University of Minnesota/TRIA	Elizabeth A. Arendt Marc Tompkins
University of Missouri	Seth Sherman

lection of PRO by any electronic media, (2) can send automated updates related to follow-up and incomplete data, and (3) can help with data analysis due to advanced data output functions.

Protocol development was conducted by a team, and clinical and radiographic assessment tools were selected. Assessment is aided by use of Web-based data collection and management for clinical outcomes and centralized radiographic evaluation. Data is collected centrally.

44.3.3 Radiographic Evaluation

Regarding radiographic evaluation, an extensive amount of time was spent in investigator meetings to develop standard radiographic methods.

Given the relative complexity of radiographic evaluation (e.g., for the measurement of anterior tibial tubercle–trochlear groove distance on MRI) in the JUPITER study, it was important that training to establish common standards for imaging evaluation was necessary. Training was performed by using a standard set of images reviewed at in-person meetings and also distributed electronically. Ultimately, concerns still remained about variability in image interpretation across sites, and part of the way through enrollment, the investigators were able to obtain sufficient funding to have images electronically sent to a central site to be interpreted by a specifically trained team of musculoskeletal radiologists. For this aspect, REDCAP was used to collect and store the data. This evolution to centralized radiographic evaluation and repository for image assessment is expected to improve consistency of this aspect of evaluation and also reduced the time commitment of individual sites. If there is a potential for significant variability in radiographic interpretation, then centralized imaging analysis at a single site is preferred.

44.3.4 Centralized Data Repository

The data from multiple sites must be collected and aggregated at a centralized site. This data management can be time-consuming and expensive; however, it is critically important to be able to have the data in a secure location that remains accessible to appropriate researchers. This typically requires financial support, which can be either provided by the sponsoring institution or external grants.

44.3.5 Funding and Grants

Once the research protocol has been established, it is often valuable to submit for research grants from various funding sources. In many cases, the initial pilot study is self- or institution-funded; however, extrapolating the study to multiple sites requires an additional level of funding. In most nonindustry-sponsored research, support for

research personnel at each individual site is typically that site's responsibility. Individual sites may have limited resources to participate in the trial, and obtaining grant support may be critical to active site enrollment. Investigators should be active in pursuit of local support as well as from larger organizations. Using data from the pilot study, the JUPITER executive group successfully submitted an application for research funding from one of the orthopedic specialty societies to support patient-reported outcome data collection. Additional grant funding from university/departmental research funding allowed for single-site radiographic review. Ultimately, it is hoped that the JUPITER experience will allow competition for NIH-funded grants that provide multi-institutional support.

Grant funding plays an important role in supporting the research project.

44.3.6 Institutional Review Board (IRB) Approval

In a clinical trial, institutional review board (IRB) ethics approval for human research must be obtained. This can often be a complex and time-consuming effort. The process is even more complex when multiple sites are involved, and data must be transmitted between different institutions. In JUPITER, the pilot site had developed an IRB-approved protocol that served as a template for IRB protocols at the other sites. This speeded up site-specific protocol development; however multisite approval resulted in delays in initiating study at several sites, for many months in some cases. In the future, the authors would consider utilizing a central IRB for the trial for as many sites as possible, which would hopefully speed gaining the appropriate ethical approval for multiple sites and decrease time to full enrollment. Recently, the NIH has released a policy on using a single IRB for multicenter trials, which may help with this process. This can be found at <https://grants.nih.gov/policy/clinical-trials/single-irb-policy-multi-site-research.htm>.

Regulatory issues (such as IRB issues and safety monitoring) can be challenging and delay site initiation. The use of a centralized IRB may be helpful but may require extensive back and forth with a primary site. Standardized protocols and procedures help with consistent and safe data collection.

Notably, audits from IRB during trials are common, and one has to organize and prepare everything such that complete transparency and responsibility could be proven at any point during the study. Recently, one of our coordinating centers had their IRB audit the entire JUPITER study. There were some omissions and some minor lapses, which have since been corrected.

An additional component of any clinical trial for publication in most major journals is registration with a clinical trial database. In the United States, www.clinicaltrials.gov is free and is the most commonly used registration.

44.3.7 Data Safety and Monitoring Board

As part of most clinical trials, a data safety and monitoring board is often required by funding agencies. The purpose is to maintain subject safety and data integrity but also to recommend cessation of the trial for ethical reasons (e.g., failure to meet enrollment or interim analysis showing dramatic differences). Protocol deviations are also evaluated. Research coordinators at key sites can help monitor compliance and are in charge of data cleaning. Periodic audits at each institution would help ensure complete data collection.

44.3.8 Standardized Operating Procedures/Training

Once a clearly defined research protocol and standardized assessment tools have been created, a manual of operating procedures (MOP) can be developed. This can help with creating

JUPITER (adapted from MARS/MOON)
Sub-Study Proposal Sheet

Based upon "Finer" approach to clinical questions described in *Designing Clinical Research* (see below).
Feasible, Interesting, Novel, Ethical, Relevant

Name: _____ Phone Number: _____
 E-Mail: _____ Date Submitted: _____

1. Study Title: _____

2. Authors/Investigators:

1.	_____
2.	_____
3.	_____

2a. Reviewers:

1.	_____
2.	_____

3. Hypotheses: _____

4. Outcome Measures:

A.	_____
B.	_____
C.	_____

5. Significance/ Previous Studies: _____

6. Data/Information Required From Coordinating Centers: _____

7. Power Analysis. Can The Cohort Answer The Question? _____

8. Statistical Analysis Required? Who Will Perform This? _____

9. Is The Current IRB Approval Adequate? Yes No UNSURE
 If No or Unsure, Please List The Needed Amendments: _____

10. Will The Study Require Additional Funding? Yes No
 If Yes, Please State The Source. _____

11. Length of Time Needed To Complete THIS Study? _____

Based upon *Designing Clinical Research: An Epidemiologic Approach* by Stephen B. Hulley, Steven R. Cummings, Warren S. Browner, Deborah G. Grady, Thomas B. Newman. Lippincott Williams & Wilkins. Third Edition, November 1, 2006.
 Adapted from Vanderbilt University Sports Medicine and MOON Forms

Fig. 44.2 JUPITER Sub-Study Proposal Sheet

standardization of enrollment and assessment. It will also help with training of research personnel. During the course of enrollment, it is not unlikely that there may be some change of research personnel at several of the sites where they may lose their research coordinator. A manual can assist in helping sites when research assistants or coordinators change. Training can be performed either in person, through documents, or by phone or online. Regularly scheduled phone or in-person meetings can keep personnel updated.

44.3.9 Research Questions

During the course of the trial or afterward, it is common to have additional research questions emerge. How to prioritize these questions and choose which ones to pursue can be difficult. It is best practice to develop criteria for the governing committee to evaluate these proposals. The FINER (feasible, interesting, novel, ethical, relevant) criteria [4] are often used to

evaluate proposed research questions. A modification of this has been used by the MOON (Multicenter Orthopedic Outcomes Network) and MARS (Multicenter ACL Revision Study) groups, and this was adopted by the JUPITER group to determine which studies to pursue (Fig. 44.2).

44.3.10 Presentation and Publication

One of the potentially more challenging aspects of multicenter trials is how results will be presented and published. It is best to address questions of authorship and publication credit and priority before study initiation and certainly before publication submission. We feel that authorship should follow International Committee of Medical Journal Editors (ICME) criteria, which is actually required by many of the premier medical journals. Authorship must meet several criteria, including significant contributions to study design, execution, assessment, and

writing and editing. Up-to-date criteria with additional detail are provided online at <http://www.icmje.org/recommendations>.

Research questions are evaluated using the FINER (feasible, interesting, novel, ethical, relevant) criteria. Presentation and publication guidelines should be developed in advance so that there is a clear understanding about authorship.

With respect to multicenter trials, authorship questions become more complex. Fortunately, several existing models for publication authorship and priority exist. Historically, many journals limited the number of named authors; however, with the advent of electronic publication and indexing, it has become easier to credit multiple authors on a publication. In many cases, papers will be published with several principal authors and the rest of the investigators credited as a group, with each individual investigator's name listed and searchable electronically. For JUPITER, authorship criteria were first discussed among the executive committee and then circulated among the larger group of investigators for comment. Criteria were modeled after the PRISM (Pediatric Research in Sports Medicine) group criteria and included assessment of active participation in data collection, as well as ICMJE criteria. To encourage multiple author participation, it was recommended that principal authors from different sites be listed on each research paper. Publications would be submitted to the group for editorial review and input as required by ICMJE.

44.4 Execution

44.4.1 Communication and Coordination

Throughout the conduct of the trial, it is critical to keep investigators and sites engaged in the research process. Too often, multicenter trials lose focus and energy since geographic distance

and limited face-to-face engagement can result in investigator focus being directed elsewhere. JUPITER has successfully addressed this with regularly scheduled monthly conference calls that include clinical research staff (such as research assistants and coordinators) as well as investigators. During these calls, critical operational updates can be provided to the group and especially the personnel that are typically performing much of the day-to-day enrollment and data collection activity. There is also time for investigators or coordinators to bring up and discuss questions or areas where further clarification is needed. Summary minutes provide valuable information that can provide updates to an existing standardized protocol.

Communication and monitoring are critical to study progress. The use of regularly scheduled meetings improves communication and consistency; transparency about site-specific trial milestones helps monitor progress and encourages continued investigator participation.

Another successful tool has been to send out frequent regular score cards indicating progress to specific trial milestones, including investigators and sites, IRB status, and their current enrollment numbers. Subject visit and follow-up compliance are additional measures to be potentially added. The score cards serve several functions. First, they update the entire group of investigators as to the current status with respect to overall enrollment in the project. It is motivating to see the progress being made across the different locations as the trial progresses. Secondly, it allows the group to identify and learn from the sites that are most successful in terms of enrollment. Finally, it can spur some friendly competition and additional engagement to increase enrollment activity.

Face-to-face group meetings are also valuable to engage the group and continue active participation. We have tried to have face-to-face meetings at major medical conferences where it is anticipated

that there will be multiple investigators available. This can be challenging since not all investigators will be at every conference, and even if investigators attend the conference, they may have other commitments that limit their availability. Some authors have suggested to address this issue, separate investigator meetings are helpful; however, this can be a significant time and expense burden.

44.4.2 Data Monitoring

Enrollment, ongoing participation, and continued follow-up need to be monitored during the course of the trial. In this way, accurate progress to milestones can be assessed and communicated. We recommend significant transparency throughout this process as there can be loss of trial participation at every step. Digital forms can significantly help in monitoring progress.

44.5 Publication

As previously noted, discussion of authorship and priority should be performed as early as possible, including in the planning phase of the study. Multiple papers typically emerge from a multicenter trial. Commonly, a methods paper is a first publication and is based primarily on the research protocol. Elements of the main paper (particularly the introduction and methods) can be written prior to obtaining complete data and analysis. Secondary studies can also be proposed prior to trial completion and evaluated as previously discussed using FINER criteria.

Multiple papers often arise from a multicenter trial. Authorship should follow guidelines. Initial components of paper writing (such as the methods section) can proceed in parallel with study recruitment. Authors should respond in a timely fashion.

The data analytics team should be notified when collection is complete so that they can

begin to work in a timely fashion. Appropriate involvement of a biostatistician in planning the study design can make the analytical work more straightforward when the data has been collected. Trying to make sense of a pile of data after the fact without appropriate preparation can be challenging.

Paper drafts should be completed promptly, and coauthors should commit to providing rapid review and comments to the drafts, hopefully within 1–2 weeks. The main author/s should assess and appropriately incorporate these comments and prepare for submission. Determination of the appropriate journal to submit to should involve the lead authors.

Following submission, it is not uncommon for high-quality journals to either reject or request significant revisions to the article. The lead author/s should take the responsibility to respond to comments and revise the article for resubmission or submission to a new journal. It is appropriate for the group to celebrate after publication!

44.6 Tips for Multicenter Trials

Multicenter trials have unique challenges in that there are additional layers of complexity due to the multiple parties involved. It is critical to have a highly motivated core group of investigators that are committed to the project. An important aspect is the inclusion of an epidemiologist/statistician in study design and planning. To simplify the process, centralization is helpful. A central IRB may help decrease time to initiation of the trial in multiple centers. Centralized image analysis can improve consistency and decrease investigator burdens. Centralized data collection is critical to a successful trial.

It is important to be aggressive in seeking out funding opportunities. Multicenter trials typically require funding of the central coordinating site and also funding of resources at each of the contributing sites. Early preparation and submission of grants can help significantly in getting the research off the ground.

Tips for Multicenter Trials

- Unique challenges due to multiple parties.
- Include epidemiologist/statistician in study design and planning.
- Centralized IRB and data collection and analysis are helpful.
- Funding can be challenging but is often needed to support the central coordinating site.
- Take advantage of others' knowledge; the MARS/MOON and PRISM multicenter trial groups were very helpful in setting up JUPITER.
- These are complex, so anticipate challenges to enrollment.
- Communication and transparency maintain consistency and encourage investigators.
- Multicenter trials have unique benefits—not only to answer research questions but also to build collegiality and collaboration with investigators from multiple institutions.

It's very helpful to take advantage of opportunities to discuss with other groups that have initiated and successfully executed multicenter trials. In orthopedic sports medicine, the MARS and MOON groups have been very helpful and have been generous with their advice. The pediatric PRISM (Pediatric Research in Sports Medicine) group has also provided models of how to address some of the questions about authorship.

One should expect that there will be difficulties in conducting the study. Sites may have difficulty with IRB approval or after beginning may lose their coordinator. It is helpful to anticipate this so build in additional sites and/or time for recruitment and enrollment.

Communication is critical throughout the process, for several reasons. It maintains the interest of geographically separated investigators. It improves the creation and conduct of the trial. Transparency with score cards regarding completion of trial milestones also is important to engage ongoing enrollment.

Finally, multicenter trials are critically important for medicine, but they are also a great way to build collegiality and friendship with investigators across multiple institutions. A critical part of medicine is shared knowledge, and working together with like-minded, interested investigators builds the community of scholars that contributes to the advancement of clinical care.

Clinical Vignette/Case Study

The JUPITER (Justifying Patellar Instability Treatment by Early Results) study is an example of how a multicenter trial is initiated, developed, and executed. After an initial pilot study at a single institution, a small group of investigators was established to develop a protocol. Additional investigators were invited to participate, and the protocol was further developed, including the decision to centralize data collection and analysis. Input was received from investigators from other multicenter studies, such as MOON, MARS, and PRISM. It was decided to use the FINER criteria for evaluation of proposed research questions, and authorship guidelines were developed as well. Some challenges were encountered with obtaining local IRB approval at several sites, and others had difficulty with research coordinator recruitment, but ultimately enrollment goals were able to be achieved, and longitudinal data is in the process of being gathered.

Take-Home Messages

- Multicenter trials have unique advantages in obtaining large numbers and increased generalizability of results but have coordination challenges.
- Careful research design, including an epidemiologist/statistician, is critical.
- Centralized data storage and analysis can improve consistency.
- Other orthopedic multicenter trials and their investigators are a valuable resource.

- Challenges are likely to arise during the IRB process and conducting the study, so additional leeway should be included for possibly delays.
- Communication between investigators is critical.
- Multicenter trials can build collegial relationships and further collaborations.

References

1. Askenberger M, Janarv PM, Finnbogason T, Arendt EA. Morphology and anatomic patellar instability risk factors in first-time traumatic lateral patellar dislocations: a prospective magnetic resonance imaging study in skeletally immature children. *Am J Sports Med.* 2017;45(1):50–8.
2. Christensen TC, Sanders TL, Pareek A, Mohan R, Dahm DL, Krych AJ. Risk factors and time to recurrent ipsilateral and contralateral patellar dislocations. *Am J Sports Med.* 2017;45(9):2105–10.
3. Chung KC, Song JW, Group WS. A guide to organizing a multicenter clinical trial. *Plast Reconstr Surg.* 2010;126(2):515–23.
4. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Designing clinical research: an epidemiologic approach. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2006.
5. Jaquith BP, Parikh SN. Predictors of recurrent patellar instability in children and adolescents after first-time dislocation. *J Pediatr Orthop.* 2017;37(7):484–90.
6. Koh JL, Stewart C. Patellar instability. *Orthop Clin North Am.* 2015;46(1):147–57.
7. Krych AJ, O’Malley MP, Johnson NR, Mohan R, Hewett TE, Stuart MJ, et al. Functional testing and return to sport following stabilization surgery for recurrent lateral patellar instability in competitive athletes. *Knee Surg Sports Traumatol Arthrosc.* 2018;26(3):711–8.
8. Sanders TL, Pareek A, Hewett TE, Stuart MJ, Dahm DL, Krych AJ. High rate of recurrent patellar dislocation in skeletally immature patients: a long-term population-based study. *Knee Surg Sports Traumatol Arthrosc.* 2018;26(4):1037–43.
9. Tompkins MA, Arendt EA. Patellar instability factors in isolated medial patellofemoral ligament reconstructions—what does the literature tell us? A systematic review. *Am J Sports Med.* 2015;43(9):2318–27.
10. Vavken P, Wimmer MD, Camathias C, Quidde J, Valderrabano V, Pagenstert G. Treating patella instability in skeletally immature patients. *Arthroscopy.* 2013;29(8):1410–22.
11. Weber AE, Nathani A, Dines JS, Allen AA, Shubin-Stein BE, Arendt EA, et al. An algorithmic approach to the management of recurrent lateral patellar dislocation. *J Bone Joint Surg Am.* 2016;98(5):417–27.



How to Organise an International Register in Compliance with the European GDPR: Walking in the Footsteps of the PAMI Project (Paediatric ACL Monitoring Initiative)

Daniel Theisen, Håvard Moksnes, Cyrille Hardy, Lars Engebretsen, and Romain Seil

45.1 Introduction

Consideration of ethical, legal and regulatory norms and standards for medical research involving human subjects has been emphasised for over 50 years. More recently, the protection of personal data has taken centre stage in the light of rapid technological developments and globalisation that have transformed human activities to an unprecedented scale. The facilitated cross-border flow of personal data has prompted the European Union (EU) to implement a strong and coherent

legal framework to ensure adequate protection of personal data, the General Data Protection Regulation 2016/679, termed hereafter GDPR.¹ The GDPR applies to all organisations collecting, processing and holding personal data of data subjects residing in the EU, independently of whether the organisation is located within or outside of the EU. The GDPR also applies to research organisations and to the vast majority of their scientific activities. Researchers active in the field of clinical orthopaedics need to be aware of how this new EU regulation impacts the organisation of their research projects.

The aim of this chapter is to highlight some important key points to be considered when implementing clinical orthopaedic research under the new European GDPR in general and setting up an international register in particular. To this end, this contribution is organised in two parts. The first illustrates GDPR requirements most relevant for the context of clinical orthopaedic research. The second presents the organisational structure of PAMI, the ESSKA Paediatric

D. Theisen (✉) · C. Hardy
Sports Medicine Research Laboratory, Luxembourg
Institute of Health, Luxembourg, Luxembourg
e-mail: daniel.theisen@lih.lu

H. Moksnes
Oslo Sports Trauma Research Center, Norwegian
School of Sport Sciences, Oslo, Norway

L. Engebretsen
Oslo Sports Trauma Research Center, Norwegian
School of Sport Sciences, Oslo, Norway

Division of Orthopedics, Oslo University Hospital,
University of Oslo, Oslo, Norway

R. Seil
Sports Medicine Research Laboratory, Luxembourg
Institute of Health, Luxembourg, Luxembourg

Department of Orthopaedic Surgery, Centre
Hospitalier Luxembourg, Luxembourg, Luxembourg

¹Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal L 119, 4.5.2016, p. 1–88.

ACL Monitoring Initiative, as an illustrative example of an international register recently implemented in Luxembourg, Europe. The reader is cautioned not to consider the information provided in this chapter to be exhaustive in any way and is strongly recommended to take legal counsel before setting up his/her research project.

Fact Box 45.1: A New European Law on Data Protection

The European Union (EU) General Data Protection Regulation (GDPR) applies to all organisations that hold or process personal data of data subjects residing in the EU (whether they are EU citizens or not), independently of the organisation's location. It also applies to research organisations and concerns the vast majority of scientific activities. GDPR is a binding legislative act that must be applied in its entirety across the EU without requiring national enabling legislation. It has come into force on May 25, 2018. Any processing under way before that date should be made compliant with the GDPR within a period of 2 years.

45.2 Background and Aim of GDPR

The general aim of the GDPR is to guarantee a consistent and high protection of the rights and freedoms of natural persons within the EU, while at the same time facilitating the flows of personal data within the Union. These aspects are of particular interest in international clinical orthopaedic research involving human participants, where most of the collected data is related to health and thus by their nature considered as “sensitive”.

Before May 2018, data protection laws in the EU were implementations by its member states and members of EØS (Norway) of the Directive 95/46/EC on “the protection of individuals with regard to the processing of personal data and on

the free movement of such data”.² Discrepancies in the specific national laws of EU member states resulted in a lack of harmonisation, which is why the EU adopted the GDPR 2016/679 in April 2016. GDPR has come into force on May 25, 2018, thus replacing the previous directive. Being a regulation, GDPR is a binding legislative act that must be applied in its entirety across the EU without requiring national enabling legislation.

45.3 Some Important Definitions

In its article 4, the GDPR specifies a number of definitions that the researcher in clinical orthopaedics should be familiar with. The most important ones are presented hereafter.

Personal data—any information, electronic or not, that relates to an identified or identifiable natural person, i.e. a “data subject”. Examples include the person’s name, address, title and contact details but also a study identification number, IP address or a social security number. In other words, compliance to the GDPR is not required if no link whatsoever can be established between the processed data and the person they belong to. This is, however, very rarely the case. All data pertaining to a person’s health (clinical outcome questionnaire, treatment, biological sample, biometrical measurement, etc.) is considered sensitive. The processing of sensitive data requires that the concerned data subject has provided explicit consent, as will be further explained below.

Processing—any operation performed on personal data, automated or not, such as collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.

²Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal L 281, 23/11/1995 P. 0031–0050.

Pseudonymisation—the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person.

Data controller—the natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data. In general, the principal investigator bears data controller responsibility.

Data processor—a natural or legal person, public authority, agency or other body which processes personal data on behalf of the controller.

Consent of the data subject—any freely given, specific, informed and unambiguous indication of the data subject’s wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.

45.4 To Be or Not to Be GDPR Compliant

Clinical Vignette 1: Drs Frank and Stein Want to Team Up in an International Research Project

Dr Frank is an orthopaedic surgeon and researcher from Europe who has been invited by his colleague Dr Stein from Canada to join an international, multicentre research project on meniscus injuries in competitive high jumpers. The leadership lies with the institution of Dr Stein who is the principal investigator of the project. Dr Frank has received a copy of the study protocol including the patient information sheet and the patient consent form. Reading through these documents, he realises that there is some ambiguity as to where the responsibilities lie regarding the

handling of personal data, in particular as to the identification of the data controller and data processor. The dataset that is supposed to be collected entails sufficient detail so the patient’s identity could be traced even if the data were pseudonymised. Before joining the research project, Dr Frank decides to contact the data protection officer from his legal department to help him design a new set of patient documents that is compliant with the EU GDPR.

Both the data controller and the data processor are responsible to ensure compliance to the GDPR when processing personal data and must be able to prove so. In particular, they must ensure that “appropriate technical and organisational measures” are implemented and can be held accountable by regulatory authorities. Non-compliance to the GDPR may expose both parties to financial risks (administrative fine up to 4% of the corporate annual turnover or € 20 Mio), in addition to business and corporate image hazards.

Any data processing under way before the date on which GDPR came into force should be made compliant within a period of 2 years. If data processing in relation to a scientific study is based on consent compliant with Directive 95/46/EC, it is not necessary for the study participant to give his or her consent again, provided that the manner in which the consent has been given complies with the GDPR.

To be compliant to the GDPR within the framework of a scientific activity, a controller or processor must:

- Maintain a GDPR registry of his activities
- Perform a data protection impact assessment (DPIA) for each new activity that handles sensitive data
- Take organisational and technical measures to mitigate the remaining risks and to manage the data subjects’ rights (e.g. pseudonymisation, data minimisation, storage limitation,

protection against accidental loss, destruction or damage, extent of data accessibility and processing, etc.)

- Adapt their agreements with stakeholders, subcontractors and data providers
- Adapt their study participant information and consent forms

Additional principles need to be considered when setting up a scientific research project involving human participants, although the following elements are in no way to be considered exhaustive.³ Study participants must be fully informed about the study purposes in an intelligible and easily accessible way, using plain and clear language, in accordance with the ICH GCP. The data controller and data processor should be clearly identified in the patient information form, and the data categories and legal framework should be described. The purpose of personal data processing should be explained, although this is not always possible at the moment of data collection. Therefore, study participants should be given the opportunity to provide their consent to certain areas of research or parts of research projects to the extent allowed by the intended purpose.

Explicit consent to participate in a study must be provided by the patient prior to participation for data processing to be lawful, and the data controller must be able to demonstrate that consent has been given. Participants must also have the possibility to withdraw consent as easily and at any moment in time, a request that must be met by the data controller. Other participant rights include the right of data access, rectification or erasure (known as “the right to be forgotten”), the right to restriction of processing and the right to data portability. The latter concerns the right of a data subject to receive his or her personal data provided to a controller in a structured and commonly readable format, so as to have the possibility to transmit this data to another controller.

³The reader is kindly referred to the chapter 14 by Mrs C. Mouton et al. in this book.

Fact Box 45.2: Some Key Points of the GDPR for Scientific Research

- Applicability of its jurisdiction to research institutions outside the European Union
- Accountability of both the data controllers and the data processors
- Improved rights for study participants, including the right to access their data, the right “to be forgotten”, the right to receive data breach notification and the right to data portability
- Principles of “privacy by default” and “privacy by design”
- Increased control and sanction regime

45.5 The Paediatric ACL Monitoring Initiative (PAMI)

Instability and functional impairments following anterior cruciate ligament (ACL) tears in skeletally immature patients represent a serious problem, which has received increasing recognition over the past years. There have been a rising number of publications on the treatment of ACL injuries in the skeletally immature population over the past decade [1, 2, 4, 5, 11, 14, 15]. Intrasubstance ACL ruptures are most worrisome due to the serious long-term health effects of potential early-onset osteoarthritis [9]. Furthermore, the open growth plates on both sides of the knee joint warrant particular caution before surgical interventions involving ACL reconstructions [3, 8]. Treatment algorithms for ACL ruptures in skeletally immature children are different around the world [10], and the only consensus currently available is that the treatment of these injuries is controversial [6, 7, 12].

To address this problem, the “Paediatric Anterior Cruciate Ligament Monitoring Initiative” (PAMI) was recently initiated [10, 13]. The main purpose of this project is to collect and analyse data from orthopaedic surgeons who are treating children and adolescents with ACL injury using a dedicated data collection system.

The ultimate goals of the PAMI project are:

1. To describe current treatment options following a paediatric ACL injury
2. To analyse the associated short-, medium- and long-term clinical outcome
3. To extend the evidence base on optimal treatment choices
4. To propose international treatment guidelines

To reach these aims, an electronic data collection platform has been created, the PAMI database, to store systematic and standardised information relevant to the problematic of paediatric ACL injuries in different European countries. Based on a long-term patient follow-up, this project will provide important insights into the current outcomes of ACL injured knees in children and will allow to discriminate those patients needing operative treatment from those who benefit most from a nonoperative treatment. Furthermore, large-scale objective outcome data will provide the knowledge base necessary for a first-time-ever proposal of international treatment guidelines.

45.6 PAMI Stakeholders

The PAMI project is initiated, promoted and financially supported by the international umbrella organisation ESSKA (European Society of Sports Traumatology, Knee Surgery and Arthroscopy—www.esska.org). Through a dedicated steering committee, ESSKA acts as the overall project coordinator and thus bears data controller responsibility.

Furthermore, the project is undertaken in collaboration with the Sports Medicine Research Laboratory of the Luxembourg Institute of Health (LIH—www.lih.lu) with extensive scientific and technical experience regarding the project objectives. The LIH is responsible for the development, deployment, maintenance and security of the PAMI database and Web application and acts as the main data processor. Data communications to and from the PAMI platform are managed via a secured communication protocol allowing for

authentication of the visited website as well as confidentiality and integrity of the data exchanged through encryption. To maximise data security, a two-factor authentication (2FA) solution has been implemented using SMS (short message service), thus adding a second level of authentication to a regular account login by project participants.

Finally, partner institutions/hospitals from different European countries providing treatment to the patient group of interest can engage in this project. They provide data on current surgical and non-surgical treatments, follow-up treatments and clinical outcome using the dedicated Web application and act as local data administrators and data processors. Only pseudonymised patient information is uploaded into the PAMI database, to ensure maximal data protection and avoid legal issues related to data transfer between different European countries. As a consequence, each site coordinator manages a correspondence table between patient ID information and a primary key generated by the system. Figure 45.1 below depicts the data flow of the upload of pseudonymised patient data.

45.7 Patient Data Processing in the PAMI Project

Clinical Vignette 2: A New Patient Joining the PAMI Project

Nancy is a 14-year-old handball player who tore the ACL of her left knee during an international tournament 6 weeks ago. During the last medical visit that she attended with her parents, she was recommended to undergo ACL reconstruction. The hospital where Nancy is treated is a partner of the PAMI research project, and her doctor explained the importance of this research and what Nancy's participation would entail. After reading the patient's information form and asking for some additional

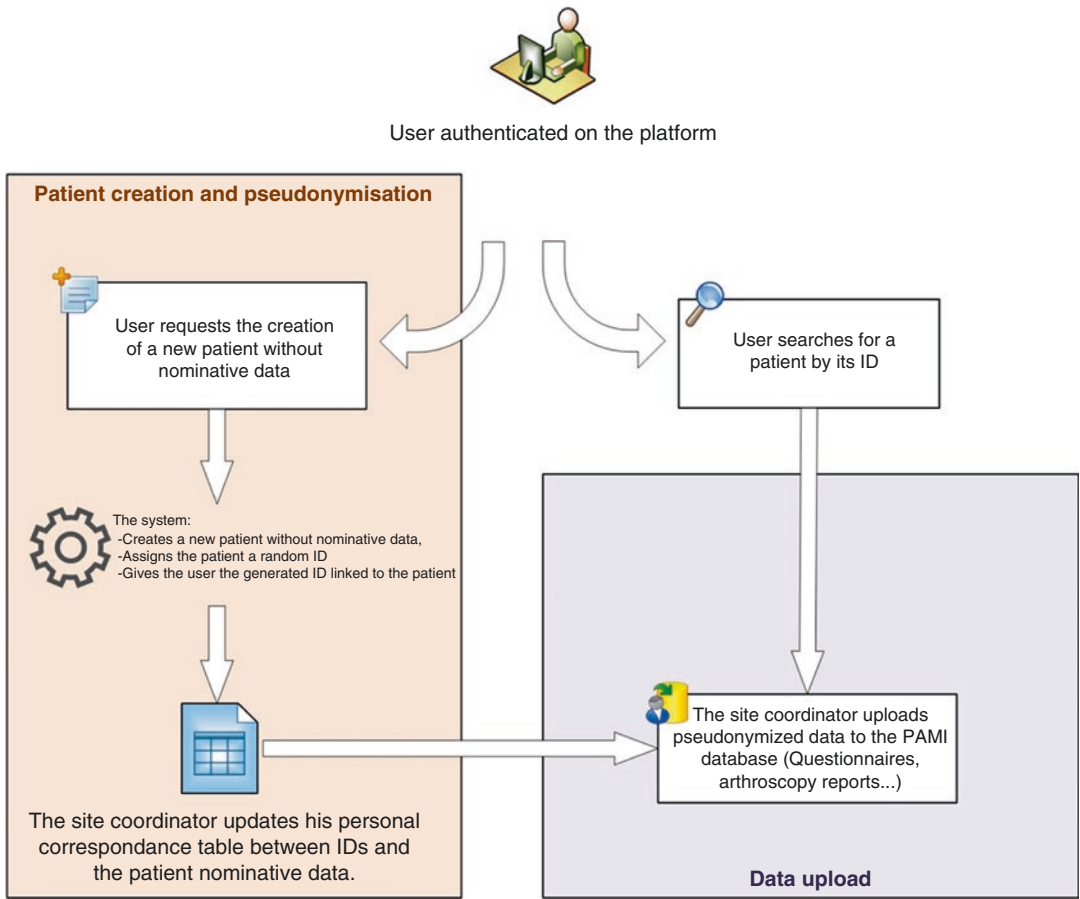


Fig. 45.1 Upload dataflow within the PAMI project

explanations, her parents agreed to provide written informed consent for her participation in the research project. Data regarding Nancy’s condition and the surgical procedures used will be uploaded into the PAMI database by the local data processor. In addition, Nancy will be asked to respond to two questionnaires on a yearly basis, one on her knee function (International Knee Documentation Committee Subjective Knee Form in Children, Pedi-IKDC) and one on her habitual physical and sports activities (Paediatric Activity Rating Scale).

Institutions/hospitals participating in the PAMI project are responsible for patient recruitment, patient information and collection of written consent for participation. Patients involved in the project will be the owners of their data. Partner institutions will act as their patients’ data administrators. Every patient has the right to access their own data as per request to the participating institution/hospital. A request can also be directly addressed by the patient to the data controller (ESSKA). Both stakeholders (data controller and local data processor) are clearly identified on the patient information form.

Institutions/hospitals willing to participate in the PAMI project must designate a single site coordinator (natural person) acting on their behalf within the PAMI project: this is the only

local person with access rights to the PAMI database. Since only pseudonymised data are collected within the PAMI project, each site coordinator manages a correspondence table between patient ID information and a primary key generated by the system. Site coordinators are also responsible for the long-term follow-up of patients by sending them annual electronic questionnaires on patient-reported clinical outcomes and physical activity/sport participation.

Prior to participation in the PAMI project, institutions/hospitals have to seek for ethics clearance to their local or national ethics committee when applicable, in accordance with their national laws and regulations. Formal proof of ethics clearance must be provided to the data controller and is a prerequisite for participation in the PAMI project. Written consent will be sought by the participating institutions from both legal tutors for patients under the age of majority, which may be different in each country; when the patient reaches the age of majority, he/she will be contacted and asked to provide written consent himself/herself. It is the responsibility of each participating partner to obtain written consent of their participants and to archive all original hard copies. The PAMI steering committee, acting on behalf of the data controller, reserves the right to perform on-site audits of participating institutions/hospitals. The date of provided consent or, if applicable, the date of consent withdrawal will be stored in the patient's record within the PAMI database and be visible to the data controller. Every participating institution will retain the right to access the pseudonymised data concerning their patients.

Ideally, the PAMI database will enable very long-term follow-up of (originally) skeletally immature children and adolescents who have suffered an ACL injury. The foreseen timeframe of follow-up will be 30 years for each patient. At the end of the 30-year follow-up, the data will be stored for an additional 10 years before being deleted. Data will be recorded for as long as the patients are willing to comply with the annual data collection.

Take-Home Message

- A new General Data Protection Regulation of the European Union has been enforced since May 2018.
- Research organisations processing personal data of residents of the EU must comply with this regulation.
- Researchers in clinical orthopaedics should take into account these principles as part of risk management of their research projects.
- The present chapter presents the most important aspects that need to be considered, as well as an illustrative example of an international register on paediatric ACL treatments.
- Researchers are strongly advised to take legal counsel with a qualified data protection officer as soon as the project planning phase starts.

References

1. Ardern CL, et al. 2018 International Olympic Committee consensus statement on prevention, diagnosis and management of paediatric anterior cruciate ligament (ACL) injuries. *Knee Surg Sports Traumatol Arthrosc.* 2018;26:989–1010. <https://doi.org/10.1007/s00167-018-4865-y>.
2. Astur DC, Cachoeira CM, da Silva Vieira T, Debieux P, Kaleka CC, Cohen M. Increased incidence of anterior cruciate ligament revision surgery in paediatric versus adult population. *Knee Surg Sports Traumatol Arthrosc.* 2018;26:1362–6. <https://doi.org/10.1007/s00167-017-4727-z>.
3. Caine D, DiFiori J, Maffulli N. Physeal injuries in children's and youth sports: reasons for concern? *Br J Sports Med.* 2006;40:749–60. <https://doi.org/10.1136/bjism.2005.017822>.
4. Calvo R, et al. Transphyseal anterior cruciate ligament reconstruction in patients with open physes: 10-year follow-up study. *Am J Sports Med.* 2015;43:289–94. <https://doi.org/10.1177/0363546514557939>.
5. Dekker TJ, Godin JA, Dale KM, Garrett WE, Taylor DC, Riboh JC. Return to sport after pediatric anterior cruciate ligament reconstruction and its effect on subsequent anterior cruciate ligament injury. *J Bone Joint Surg Am.* 2017;99:897–904. <https://doi.org/10.2106/JBJS.16.00758>.
6. Frosch KH, et al. Outcomes and risks of operative treatment of rupture of the anterior cruciate ligament in children and adolescents. *Arthroscopy.* 2010;26:1539–50. <https://doi.org/10.1016/j.arthro.2010.04.077>.

7. Kaeding CC, Flanigan D, Donaldson C. Surgical techniques and outcomes after anterior cruciate ligament reconstruction in preadolescent patients. *Arthroscopy*. 2010;26:1530–8. <https://doi.org/10.1016/j.arthro.2010.04.065>.
8. Kocher MS, Saxon HS, Hovis WD, Hawkins RJ. Management and complications of anterior cruciate ligament injuries in skeletally immature patients: survey of the Herodicus Society and The ACL Study Group. *J Pediatr Orthop*. 2002;22:452–7.
9. Lawrence JT, Argawal N, Ganley TJ. Degeneration of the knee joint in skeletally immature patients with a diagnosis of an anterior cruciate ligament tear: is there harm in delay of treatment? *Am J Sports Med*. 2011;39:2582–7. <https://doi.org/10.1177/0363546511420818>.
10. Moksnes H, Engebretsen L, Seil R. The ESSKA paediatric anterior cruciate ligament monitoring initiative. *Knee Surg Sports Traumatol Arthrosc*. 2016;24:680–7. <https://doi.org/10.1007/s00167-015-3746-x>.
11. Pierce TP, Issa K, Festa A, Scillia AJ, McInerney VK. Pediatric anterior cruciate ligament reconstruction. *Am J Sports Med*. 2017;45:488–94. <https://doi.org/10.1177/0363546516638079>.
12. Reider B. A matter of timing. *Am J Sports Med*. 2015;43:273–4. <https://doi.org/10.1177/0363546515570023>.
13. Seil R, Theisen D, Moksnes H, Engebretsen L. ESSKA partners and the IOC join forces to improve children ACL treatment. *Knee Surg Sports Traumatol Arthrosc*. 2018;26:983–4. <https://doi.org/10.1007/s00167-018-4887-5>.
14. Shaw L, Finch CF. Trends in pediatric and adolescent anterior cruciate ligament injuries in Victoria, Australia 2005–2015. *Int J Environ Res Public Health*. 2017;14. <https://doi.org/10.3390/ijerph14060599>.
15. Wall EJ, Ghattas PJ, Eismann EA, Myer GD, Carr P. Outcomes and complications after all-epiphyseal anterior cruciate ligament reconstruction in skeletally immature patients. *Orthop J Sports Med*. 2017;5:232596711769360. <https://doi.org/10.1177/2325967117693604>.

Part X

Helpful Further Information



Common Scales and Checklists in Sports Medicine Research

46

Alberto Grassi, Luca Macchiarola, Marco Casali,
Ilaria Cucurnia, and Stefano Zaffagnini

46.1 Introduction

Since the main purpose of clinical studies, especially randomized controlled trials (RCT), is to report or compare the effect of different treatments, the measurement methods of clinical outcomes are crucial. Therefore, during the early stage of study design, attention should be directed to choosing the appropriate outcomes and scales that evaluate a patient population. The calculation of sample size of a RCT is primarily based on the primary outcome being evaluated. When dichotomous outcomes of rare events such as failures or complications are used, extremely large sample sizes are often required. This requirement may discourage the realization of the study or require

an enormous amount of resources to reach adequate enrollment. Conversely, if continuous measures such as patient- or clinician-reported scales are chosen, a power analysis based on means and standard deviations usually provides more feasible sample size.

However, care should be used during the power analysis to ensure that it is based on a clinical score which is able to detect a real difference between the different treatments. In fact, building a clinical study or RCT on outcomes which are not completely appropriate to the study purpose, patient population, and treatment administered could compromise the utility and subsequent impact of the results. Therefore, the researcher should be very familiar with the main features of clinical scores and should also know the main characteristics of each scale in relation to the pathology or treatment that is being investigated.

Another important aspect in outcome selection is the global assessment of the patient. Traditionally, clinical outcomes in orthopedics consisted of measuring impairments such as range of motion, joint stability, strength, pain, and joint function. At times, surgeons are marginally interested in patient's global disability and mental status; however, the patient's perception of changes in health status is the most important indicator of the success of a treatment. Therefore, there are two possibilities of measuring health-related quality of life in orthopedic and sports medicine conditions. The "generic measures"

A. Grassi
IIa Clinica Ortopedica e Traumatologica, IRCCS
Istituto Ortopedico Rizzoli, Bologna, Italy

Dipartimento di Scienze Biomediche e Neuromotorie
(DIBINEM), Università di Bologna, Bologna, Italy

SIGASCOT Arthroscopy Committee, Florence, Italy
e-mail: alberto.grassi@ior.it

L. Macchiarola (✉) · M. Casali · I. Cucurnia
IIa Clinica Ortopedica e Traumatologica, IRCCS
Istituto Ortopedico Rizzoli, Bologna, Italy

S. Zaffagnini
IIa Clinica Ortopedica e Traumatologica, IRCCS
Istituto Ortopedico Rizzoli, Bologna, Italy

Dipartimento di Scienze Biomediche e Neuromotorie
(DIBINEM), Università di Bologna, Bologna, Italy
e-mail: stefano.zaffagnini@unibo.it

pertain the overall health of the patient, including physical, mental, and social well-being, and offer the advantage of being able to use them to compare different diseases, severity, and interventions. However, since they represent a generic measure, their ability to detect small but important changes could be limited. On the other hand, the “disease-specific measures,” which pertain to a specific disorder treated in a patient, measure the physical, mental, and social aspects of health affected by the specific disorder. Therefore, they are able to detect small but important changes but have a limited value in comparison of health status across different diseases. For the aforementioned reasons, a complete picture of treatment effect on a patient could be provided only with the assessment through a “disease-specific measure” in combination with a “generic measure.”

Fact Box 46.1

Too often surgeons are poorly interested in patient’s global health self-perception and mental status; however, the patient’s perception of changes in health status and disability is the main indicator of the success of a treatment.

46.2 General Scale Characteristics

The main characteristics and features of clinical scales that should be known to choose the appropriate outcome measure are the following [54].

Construct validity: is defined as the ability of an instrument to measure what it is supposed to measure. It depends on how the items that make up the scale include all relevant aspects of the pathology or disability that is measured. The convergent validity indicates how the score could correlate with other scores that measure the same construct. Meanwhile, predictive validity indicates whether the score could predict a patient’s score on a measure of some related construct.

Repeatability (Test/retest reliability): is defined as the agreement between the observations on the same patients on two or more occa-

sions separated by a time interval under stable health conditions. It is considered when the raters are not involved or the raters’ effect is negligible. It could be assessed with the intraclass correlation (ICC) or the Cohen’s *K* statistics.

Intra-rater reliability: is defined as the agreement between two or more repeated score evaluation performed by a single rater. Also in this case, intraclass correlation (ICC) or the Cohen’s *K* statistics could be used.

Inter-rater reliability: is defined as the agreement between the scores obtained from two or more raters’ assessment. It measures how much consensus or heterogeneity there is in the rating given by judges. Similarly, intraclass correlation (ICC) or the Cohen’s *K* statistics is employed.

Internal consistency: is defined as the correlations between different items on the same test and measures whether several items that propose to measure the same general construct produce similar scores. It is assessed using the Cronbach’s alpha statistics.

Responsiveness to change: is defined as the ability of an instrument to detect clinically important changes between the patient’s pre-intervention and post-intervention state, assuming all other factors remain constant.

Minimal detectable change (MDC): is defined as the minimal change that falls outside the measurement error in the score of an instrument used to measure a symptom.

Minimal clinically important difference (MCID): is defined as the minimal change in the score that is meaningful for patients or that is required for the patient to feel a difference in the variable that is measured.

Standard error of measurement (SEM): It measures the range within which a score would likely fall in the case of re-measurement.

Standardized response mean (SRM): It measures the responsiveness to change and is defined as the mean change in score divided by the standard deviation of the change scores.

Floor effect: Floor effects occur when a measure’s lowest score is unable to assess a patient’s level of ability. The test is considered poor if the floor effect is >20%.

Ceiling effect: Ceiling effects occur when a measure's highest score is unable to assess a patient's level of ability. This might be particularly common for measures used over multiple occasions. The test is considered poor if the ceiling effect is $>20\%$.

46.3 Measures of Shoulder Function

There are many instruments that measure symptom and function of the shoulder and some that evaluate both the glenohumeral joint and the whole upper limb. The most widespread and best tested is the disabilities of the arm, shoulder, and hand questionnaire (DASH). Also, the shoulder pain and disability index (SPADI), the Constant-Murley score (CMS), and the American shoulder and elbow surgeons (ASES) questionnaire, which are more specific for shoulder pathologies, are extensively employed. The simple shoulder test (SST), the shoulder disability questionnaire (SDQ), the Oxford shoulder score, and the West Ontario shoulder instability index (WOSI) complete the panorama of most common tools.

The basic psychometric characteristics, strengths, and weaknesses of the most common scales for shoulder function are described (Table 46.1).

46.3.1 Disabilities of the Arm, Shoulder, and Hand Questionnaire (DASH)

The DASH is a patient-completed scale which includes 30 items regarding symptoms, pain, physical function, and social function [58]. The 11-item QuickDASH short version is also available [5]. The DASH is the best tool for the comprehensive assessment of upper extremity conditions, since it is easy to apply, analyze, and interpret; moreover it is good for research purposes in various upper extremity conditions and has a good correlation with SPADI, HAQ, CMS, ASES, and EQ-5D with Pearson's or Spearman's

test. It is particularly useful when polyarticular conditions should be evaluated or symptoms and function of the entire upper extremity are investigated. It is also useful in all elbow and hand conditions. However, the DASH is region specific and not joint specific; therefore specificity and responsiveness are lower than those of unique shoulder-specific tools [3].

Conditions: Any or multiple disorders of upper extremity, in particular painful conditions including: rheumatoid arthritis, multiple sclerosis, adhesive capsulitis, shoulder impingement and tendinitis, proximal humerus fracture, distal radius fractures, hand osteoarthritis or fractures, arthroscopic acromioplasty.

46.3.2 Shoulder Pain and Disability Index (SPADI)

The SPADI is a patient-completed scale that includes 13 items regarding symptoms and pain, scored on a VAS/NRS scale [94]. It is one of the most representative shoulder instruments and has been tested in numerous settings; moreover, it is easy to administer, understand, and complete. It has a good correlation with the DASH, ASES, and CMS. One possible weakness in construct validity could be that only one item assesses overhead work [3].

Conditions: Any disorder of the shoulder joint, particularly adhesive capsulitis, rotator cuff pathologies.

46.3.3 American Shoulder and Elbow Surgeons Society Shoulder Assessment Form (ASES)

The ASES is a patient self-evaluation scale of 11 items evaluation pain and function, which is integrated with a clinician-dependent part [92]. It has good reliability, construct validity, and responsiveness. However, it uses different type of scales (binary, Likert, VAS), and the clinician part could be time-consuming. It has been developed to be applied to all shoulder patients regardless of the diagnosis, since it evaluates also activities of

Table 46.1 Measures of shoulder function

Questionnaire	Items	Options	Administration	Range	Cutoffs	Time compilation (min)	Time calculation (min)	Internal consistency	Test-retest	SRM	MDC	MCID
DASH ^a	30	Likert (5)	Patient	0 (best) to 100 (worst)	Yes	4	10	0.92–0.98	0.93–0.98	0.43–1.2	7.9–14.8	10.2
SPADJ ^a	13	VAS/NRS	Patient	0 (worst) to 10 (best)	No	5	2	0.86–0.96	0.84–0.95	1.23–1.81	13.2–21.5	13.2–23.1
ASES ^a	11	Mix	Patient	0 (worst) to 100 (best)	No	3	>8	0.61–0.96	0.84–0.9	1.42–1.81	11.2	6.4–16.9
CMS ^a	8	Likert (3–10)	Patient + clinician	0 (worst) to 100 (best)	Yes	5	NA	0.60	0.80–0.96	0.59–2.09	NR	NR
SST ^b	12	Dichotomous	Patient	0 (worst) to 12 (best)	No	2	5	0.85	0.97–0.99	0.63–1.94	NR	2.05–2.33
OSS ^a	12	Likert (5)	Patient	12 (best) to 60 (worst)	No	3	5	0.94	0.98	1.10–1.14	NR	NR
UCLA ^b	5	Likert (2–6)	Patient + clinician	0 (worst) to 35 (best)	Yes	5	5	0.93–0.95	NR	0.15–0.90	NR	NR
WOSI ^a	21	VAS/NRS	Patient	0 (best) to 100 (worst)	No	3	6	0.88–0.96	0.87–0.98	0.93–1.40	NR	NR
WOOS ^c	19	VAS	Patient	0 (worst) to 100 (best)	No	10	5	NR	0.96	1.91	NR	NR
WORC ^d	20	VAS	Patient	0 (worst) to 100 (best)	No	10	5	NR	0.96	NR	NR	11.7
OSIQ ^e	12	Likert (5)	Patient	0 (worst) to 48 (best)	Yes	5	5	0.88	0.87	NR	9	NR

Note: DASH (disabilities of the arm, shoulder, and hand questionnaire), SPADI (shoulder pain and disability index), ASES (American shoulder and elbow surgeons society shoulder assessment form), CMS (Constant-Murley score), SST (simple shoulder test), OSS (Oxford shoulder score), UCLA (University of California at Los Angeles shoulder score), WOSI (Western Ontario shoulder instability index), WOOS (Western Ontario osteoarthritis of the shoulder index), WORC (Western Ontario rotator cuff index), OSIQ (Oxford shoulder instability questionnaire)

^aAngst F, Schwytzer HK, Aeschlimann A, Simmen BR, Goldhahn J. Measures of adult shoulder function. *Arthritis Care Res* (Hoboken). 2011 Nov;63 Suppl 11:S174–88

^bAmstutz HC, Sew Hoy AL, Clarke IC. UCLA anatomic total shoulder arthroplasty. *Clin Orthop Relat Res*. 1981 Mar-Apr;(155):7–20

^cLo IK, Griffin S, Kirkley A. The development of a disease-specific quality of life measurement tool for osteoarthritis of the shoulder: The Western Ontario Osteoarthritis of the Shoulder (WOOS) index. *Osteoarthritis Cartilage*. 2001 Nov;9(8):771–8

^dKirkley A, Griffin S, Dainty K. Scoring systems for the functional assessment of the shoulder. *Arthroscopy*. 2003 Dec;19(10):1109–20

^evan der Linde JA, van Kampen DA, van Beers LW, van Deursen DF, Terwee CB, Willems WJ. The Oxford Shoulder Instability Score; validation in Dutch and first-time assessment of its smallest detectable change. *J Orthop Surg Res*. 2015 Sep 17;10:146

daily living. It has a good correlation with both SPADI and DASH questionnaires [3].

Conditions: Any disorder of the shoulder joint, particularly rotator cuff disease, shoulder impingement, shoulder arthritis, calcific tendonitis.

46.3.4 Constant-Murley Score (CMS)

The CMS is both patient- and clinician-reported score which includes eight items regarding pain, ADLs, mobility, and strength. It is a method to record individual parameters, providing an overall clinical functional assessment, irrespective of diagnosis or radiographic abnormalities [22, 23]. Based in the difference with the abnormal side, the indexed shoulder could be graded as excellent (<11), good (11–20), fair (21–30), or poor (<30). Despite the CMS is highly accepted throughout the clinical community, there are several limitations to its use due to the low inter-tester reliability, non-standardized measurement of strength, and only few items evaluating pain and ADL. It is useful for measurement protocols but does not provide an adequate self-assessment of patient pain and function. It has a good correlation with ASES, DASH, and SPADI [3].

Conditions: Mainly rotator cuff-related disorders, impingement, degenerative or inflammatory pathologies, instability, osteoarthritis.

46.3.5 Simple Shoulder Test (SST)

The SST is a patient-reported score which includes 12 dichotomous (yes/no) items regarding pain, strength, and range of motion [71]. It assesses the functional disability of the shoulder in a very simple and short manner; however, due to the binary response option, its use as a comprehensive measure of outcomes could be questioned. It has a good correlation with the SPADI, ASES, DASH, and CMS scores [3].

Conditions: General shoulder injuries and rotator cuff pathology.

46.3.6 Oxford Shoulder Score (OSS)

The OSS is a patient-reported scale of 12 items evaluating pain and daily function [32, 35]. It provides a self-assessment of shoulder pain and function. It is short and easy to complete but not frequently used in the current literature. Correlation with SPADI, DASH, and CMS is good [3].

Conditions: Degenerative and inflammatory shoulder conditions, subacromial impingement, rotator cuff, osteoarthritis, and proximal humerus fractures.

46.3.7 UCLA Shoulder Score

The UCLA (University of California at Los Angeles) shoulder score is both a five-item patient- and clinician-reported scale which evaluates pain, function, ROM, strength, and patient's satisfaction [2]. Despite being one of the earliest available shoulder outcome measures, it has not formally been validated. It is simple and fast but requires physician manual evaluation; for this reason, it could result in a poor validity or responsiveness, which does not make it ideal for research setting. The UCLA has a good correlation with the DASH, SPADI, and SF-36 and could be dichotomized as good/excellent (>27) or fair/poor (<27) [61].

Conditions: Common shoulder pathologies.

46.3.8 Western Ontario Shoulder Instability Index (WOSI)

The WOSI is a 21-item patient-reported scale that evaluates physical symptoms, pain, sport, work, lifestyle, and emotions related to shoulder instability [62, 63]. It has been developed to assess disease-specific quality of life patients with symptomatic shoulder instability. It has the advantage of being specific for this condition, but due to lack of testing data, caution is necessary at individual patient level. It has a good correlation with the VAS for function and the DASH score [3].

Conditions: Shoulder instability.

46.3.9 Western Ontario Osteoarthritis of the Shoulder Index (WOOS)

The WOOS is a 19-item patient-reported questionnaire that evaluates the area of pain, physical symptoms, sports and work, lifestyle function, and emotional function [72]. Its form as 100-mm VAS makes it an easy, fast, and reliable questionnaire; however, it is specific for degenerative pathologies, especially osteoarthritis. Its multiple domains regarding both function and psychologic aspects make the WOOS a versatile and complete scale. In fact, it contains many items rarely investigated by other shoulder questionnaires. It has a moderate correlation with the Constant-Murley and UCLA scores [61].

Conditions: Osteoarthritis of the shoulder.

46.3.10 Western Ontario Rotator Cuff Index (WORC)

The WORC is a 20-item patient-reported score that evaluates symptoms, sport, work, emotion, and social function [60]. It is easy and rapid to administer, since it is composed of 100-mm VAS items, but it is disease-specific since it has been developed to evaluate rotator cuff-related quality of life.

It has a good correlation with the ASES and UCLA scores [61].

Conditions: Rotator cuff pathology treated surgically and conservatively.

46.3.11 Oxford Shoulder Instability Questionnaire (OSIQ)

The OSIQ is a 12-item patient-reported questionnaire that explores the impact of shoulder instability on work, sport and social life, its psychological repercussion, the quality of life, and the pain [33, 35]. It is specifically designed for glenohumeral dislocation and shoulder instability. However, it has a good correlation with both DASH and WOSI. Based on the obtained value, function could be graded as

excellent (40–48), good (30–39), fair (20–29), or poor (0–19) [108].

Conditions: Surgery or physiotherapy for shoulder instability. Shoulder instability.

46.4 Measures of Elbow, Wrist, and Hand Function

Elbow, wrist, and hand function represent a complex dimension to evaluate. Especially for elbow, physical examination and objective evaluation of ROM and stiffness are important characteristics to assess the joint function, patient's satisfaction, and normal or pathologic conditions. Therefore, clinical scores often require a clinician-reported items that increase the precision of the evaluation, but on the other side, they reduce the reliability and make them time-consuming as well.

The basic psychometric characteristics, strength, and weakness of the most common scales for elbow, wrist, and hand function are described (Table 46.2).

46.4.1 Mayo Elbow Performance Score (MEPS)

The MEPS is both a patient- and clinician-reported score. It includes four Likert-scale items evaluating mostly pain and motion, stability, and function [82]. It is correlated with other elbow measures for raw scores rather than categorical ranks and requires clinician objective evaluation of the patient, which could lengthen its application [16].

Conditions: General elbow disorders, rheumatoid arthritis, synovectomy.

46.4.2 Oxford Elbow Score (OES)

The OES is a 12-item patient-reported score. It includes 12 Likert-scale items evaluating elbow function, pain, and the psychological aspects [30]. It is easy and simple to administer to patients; however, it lacks objective evaluation of clinical outcomes. It has a good correlation with DASH, Mayo elbow score, and SF-36 [73].

Conditions: General elbow disorders.

Table 46.2 Measures of elbow, wrist, hand function

Questionnaire	Items	Options	Administration	Range	Cutoffs	Time compilation (min)	Time calculation (min)	Internal consistency	Test-retest	SRM	MDC	MCID
MEPS ^a	4	Likert (3–4)	Patient + clinician	0 (worst) to 100 (best)	No	5	3	NR	0.89	NR	11.3	15
OES ^b	12	Likert (5)	Patient	0 (worst) to 100 (best)	No	3	3	0.8–0.90	0.87	0.46–0.69	27.6	NR
ASES ^c	57	Mix	Patient + clinician	0 (worst) to 100 (best)	No	15	10	0.68–0.82	0.95	NR	NR	NR
PRTEE ^d	10	NRS	Patient	0 (best) to 100 (worst)	No	3	3	NR	0.87	2.01	NR	NR
MWS ^e	4	Likert (3–4)	Patient + clinician	0 (worst) to 100 (best)	No	5	5	NR	NR	NR	NR	NR
MHQ ^f	37	Likert (5)	Patient	0 (worst) to 100 (best)	No	15	20	0.75–0.94	0.95	0.47–1.61	NR	3–13
FIHOA ^f	10	Likert (4)	Patient	0 (best) to 30 (worst)	No	3	3	0.85–0.90	0.95	0.58–0.87	NR	NR

Note: MEPS (Mayo elbow performance score), OES (Oxford elbow score), ASES (American shoulder and elbow surgeons society shoulder assessment form), PRTEE (patient-rated tennis elbow evaluation), MWS (Mayo wrist score), MHQ (Michigan hand outcome questionnaire), FIHOA (functional index for hand osteoarthritis)

^aCelik D. Psychometric properties of the Mayo Elbow Performance Score. *Rheumatol Int.* 2015 Jun;35(6):1015–20

^bDawson J, Doll H, Boller I, Fitzpatrick R, Little C, Rees J, Jenkinson C, Carr AJ. The development and validation of a patient-reported questionnaire to assess outcomes of elbow surgery. *J Bone Joint Surg Br.* 2008 Apr;90(4):466–73

^cMacDermid JC. Outcome evaluation in patients with elbow pathology: issues in instrument development and evaluation. *J Hand Ther.* 2001 Apr-Jun;14(2):105–14

^dRompe JD, Overend TJ, MacDermid JC. Validation of the Patient-rated Tennis Elbow Evaluation Questionnaire. *J Hand Ther.* 2007 Jan-Mar;20(1):3–10

^eAmadio PC, Berquist TH, Smith DK, Ilstrup DM, Cooney WP 3rd, Linscheid RL. Scaphoid malunion. *J Hand Surg Am.* 1989 Jul;14(4):679–87

^fPoole JL. Measures of hand function. *Arthritis Care Res (Hoboken).* 2011 Nov;63 Suppl

46.4.3 American Shoulder and Elbow Surgeons Society Shoulder Assessment Form (ASES)

The ASES elbow outcome score is both a patient- and clinician-reported questionnaire that evaluates elbow pain, function, and satisfaction through 19 items and motion, stability, strength, and physical findings through 38 items [92]. This score represents a complete evaluation but requires substantial time to be complete, and pain has the highest influence in the overall score [80].

Conditions: General elbow disorders.

46.4.4 Patient-Rated Tennis Elbow Evaluation (PRTEE)

The PRTEE is a 15-item patient-reported score that evaluates forearm pain and disability in patients with lateral epicondylitis. It presents two subscales: pain and function. It is easy to complete and fast to be administered and had good correlation with NRS for pain during wrist extension and with the DASH; however, it is specific one condition [75, 97].

Conditions: Lateral epicondylitis of the elbow.

46.4.5 Mayo Wrist Score (MWS)

The MWS is both a patient- and clinician-reported score, which includes five Likert-scale items evaluating pain, function, ROM, and grip strength. It's basic but involves objective evaluation of wrist mobility and strength which could limit its usability [25]. Moreover, its reliability and consistency characteristics have not been deeply investigated. It could be graded as excellent (90–100), good (80–90), satisfactory (60–80), and poor (<60) [28].

Conditions: Originally developed for scaphoid nonunion; could be used for wrist fractures and arthritis.

46.4.6 Michigan Hand Outcome Questionnaire (MHQ)

The MHQ is a 37-item patient-reported scale that evaluates hand function, appearance, pain, and

satisfaction [19]. It appropriately measures hand function in various conditions; however, its application could be time-consuming [28].

Conditions: Hand and wrist injuries, including osteoarthritis.

46.4.7 Functional Index for Hand Osteoarthritis (FIHOA)

The FIHOA is a patient-reported scale composed of ten items including questions about using keys, cutting, lifting, buttoning, and writing, aimed to measure hand function in patients with hand osteoarthritis. It has a good correlation with the MHOQ [36].

Conditions: Hand osteoarthritis.

46.5 Measures of Hip Function

The assessment of outcomes in hip surgery is focused on patient satisfaction and the quality of life achieved, level of pain, range of motion (ROM), comorbidities, and the use of walking aids. A variety of quality of life evaluation tools have been developed that differ in their measurement techniques and in the number of domains they assess. These scores are useful not only for the normal clinical evaluation in old patients and in hip congenital disease but also to assess the outcomes after joint-preserving surgery. The ideal hip outcome measure should be one that is specific for the hip joint, possesses a generic component, and is clear and concise. Previous outcome tools were modifications of preexisting tools that evaluate chronic conditions such as osteoarthritis. Outcome measures most frequently used in clinical practice are the Harris hip score, the hip disability and osteoarthritis outcome score, the Oxford hip score, and the Lequesne index of severity for osteoarthritis of the hip. More specific scores for sport-related hip injuries were designed in the last years such as non-arthritic hip score and international hip outcome tool-33.

The basic psychometric characteristics, strength, and weakness of the most common scales for hip function are described (Table 46.3).

Table 46.3 Measures of hip function

Questionnaire	Items	Options	Administration	Range	Cutoffs	Time compilation (min)	Time calculation (min)	Internal consistency	Test-retest	SRM	MDC	MCID
HHS ^a	10	Likert (2–15)	Clinician	0 (worst) to 100 (best)	Yes	5	10	NR	0.93–0.98	2.52–2.73	NR	NR
HOOS ^b	40	Likert (5)	Patient	0 (worst) to 100 (best)	No	10–15	2–3	0.82–0.98	0.75–0.97	1.29–3.24	9.6–16.2	NR
OHS ^c	12	Likert (5)	Patient	0 (worst) to 48 (best)	Yes	2–8	5	0.84–0.93	0.84–0.93	1.12	6.11	3–5
LISOH ^a	11	Likert (2–7)	Patient + clinician	0 (best) to 24 (worst)	Yes	2–5	5	0.83–0.84	0.94	NR	NR	NR
NAHS ^c	20	Likert (5)	Patient	0 (worst) to 100 (best)	No	8–10	5	0.69–0.92	0.87–0.95	NR	10.4–12.4	NR
iHOT33 ^{d,e}	33	VAS	Patient	0 (worst) to 100 (best)	No	15	6	0.96–0.99	0.87–0.96	1.7	16.0	6.1
HAGOS ^f	37	Likert (5)	Patient	0 (worst) to 100 (best)	No	10	10	0.37–0.73	0.82–0.92	NR	02-mag	NR

Note: HHS (Harris hip score), HOOS (hip disability and osteoarthritis outcome score), OHS (Oxford hip score), LISOH (Lequesne in severity for osteoarthritis of the hip), NAHS (non-arthritis hip score), iHOT33 (international hip outcome tool-33), HAGOS (the Copenhagen hip and groin outcome score)

^aNilsdotter A, Bremander A. Measures of Hip Function and Symptoms. *Arthritis Care & Research* Vol. 63, No. S11, November 2011, pp. S200–S208

^bMartinelli N, Longo UG, et al. Cross-cultural adaptation and validation with reliability, validity, and responsiveness of the Italian version of the Oxford Hip Score in patients with hip osteoarthritis. *Qual Life Res* (2011) 20:923–929

^cRamisetty N, Kwon Y, and Mohrtdi N. Patient-reported outcome measures for hip preservation surgery. A systematic review of the literature. *Journal of Hip Preservation Surgery* Vol. 2, No. 1, pp. 15–27

^dMohrtdi NG, Griffin DR, Pedersen ME et al. The development and validation of a self-administered quality-of-life outcome measure for young, active patients with symptomatic hip disease: the international hip outcome tool (iHOT-33). *Arthroscopy* 2012; 28: 595; 605

^eKemp JL, Collins NJ, Roos EM et al. Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. *Am J Sports Med* 2013; 41: 2065–73

^fThorborg K, Holmich P, Christensen A, Petersen F, Roos EM. The Copenhagen Hip and Groin Outcome Score (HAGOS): development and validation according to the COSMIN checklist. *Br J Sports Med*. 2011;45:478–491

46.5.1 Harris Hip Score (HHS)

The HHS is a clinician-based outcome score which includes ten items that evaluates pain, function, absence of deformity, and range of motion [44]. There are two versions of the score: the original one, published in 1969, and the modified HHS (MHHS). The latter only includes pain and function components and has been widely used to evaluate outcomes in hip arthroscopy surgery. The HHS is widely used throughout the world for evaluating outcome after THR, and it has also been proven appropriate to measure outcome after surgical interventions for femoral neck fractures. It seems to be useful for short-time follow-up; moreover there are unacceptable ceiling effects that severely limit its validity. The HHS has been used in many different countries (Sweden, the Netherlands, Denmark, etc.), but there are no validated versions in other languages available. It has a good correlation with WOMAC, NHP, NAHS, and SF-36 for pain and function domain. Based on the obtained score, it could be graded as excellent (90–100), good (80–90), fair (70–80), or poor (<70) [86].

Conditions: Femoral neck fractures, osteoarthritis of hip, impingement.

46.5.2 Hip Disability and Osteoarthritis Outcome Score (HOOS)

The HOOS is a patient-reported score composed of 40 items that evaluates pain, other symptoms, function in activities of daily living, function in sport and recreation, and hip-related quality of life. It has been validated in two slightly different versions, LK1.1 and LK2.0 [65, 85]. In 2008, a five-item measure of physical function, the HOOS-PS, was published derived from the HOOS questionnaire by item response theory to elicit patients' opinions about difficulties experienced due to hip problems. The HOOS has been used in subjects with hip disability with or without hip osteoarthritis and in patients with hip osteoarthritis pre- and postoperative total hip

replacement. The HOOS is an extension of the WOMAC and is suggested to be valuable for younger and more active people due to the added subscales. It is suitable for use in research as a disease-specific questionnaire. It has a good correlation with Oxford hip score, the Lequesne index, and the VAS for pain. Based on its score, it could be graded as excellent (>41), good (34–41), fair (27–33), or poor (<27) [86].

Conditions: Osteoarthritis, general hip disorders.

46.5.3 Oxford Hip Score (OHS)

The OHS is a 12-item patient-reported outcome score regarding pain and function of the hip in relation to daily activities such as walking, dressing, and sleeping. It is designed for the assessment of joint replacement and has been used in several countries in large registry studies [31, 83]. The OHS was developed to supplement other generic outcome measures in systematic studies of hip replacement surgery with long-time follow-up. It has also been validated and used in revision hip replacement. Due to its shortness, the OHS questionnaire is feasible for surveys by mail, and it yields a high response rate and is therefore preferred for larger studies. High correlation was found between OHS and the HHS in THR patients [86].

Conditions: Osteoarthritis of hip.

46.5.4 Lequesne in Severity for Osteoarthritis of the Hip (LISOH)

The LISOH is an interview-based or reported score which includes 11 items evaluating pain, maximum distance walked, and activities of daily living [68–70]. The LISOH is available currently in several versions: interview based, self-administered, and in modified versions due to changed scoring and wording. The LISOH was developed to evaluate the severity of hip osteoarthritis in drug trials and the long-term treatment effects for hip OA and as help in decision-making regarding

the need for hip replacement. It has limited construct validity; also the convergent validity of the questionnaire has been questioned. Recommendations are to only use the LISOH for group comparisons. Based on its score, the handicap derived from hip osteoarthritis could be graded as extremely severe (>14), very severe (11–13), severe (10–8), moderate (5–7), mild (1–4), or none (0) [86].

Conditions: Osteoarthritis, the effectiveness of pharmacologic interventions, and to help with indications for surgery like THA.

46.5.5 Non-arthritic Hip Score (NAHS)

The non-arthritic hip score (NAHS) consists of 20 items distributed in four domains of pain, mechanical symptoms, functional symptoms, and activity level. It was developed for young active patients with higher demands and expectations. This is a patient-based, self-administered questionnaire that was developed as a modification of the Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC). Input from patients, surgeons, physical therapists, and epidemiologists was used in creating NAHS scoring system. The NAHS has satisfactory internal consistency in each of its four domains. But there is no further evidence about internal consistency from head-to-head comparison studies with other outcome measures. Hence, the summation score for internal consistency for NAHS is good. The summation score for test-retest reliability is excellent. Construct validity is satisfactory between the NAHS and the Harris hip score (HHS) and short form (SF)-12, respectively [18].

Conditions: All non-arthritic hip conditions.

46.5.6 International Hip Outcome Tool-33 (iHOT-33)

These 33 questions were formulated into a self-administered questionnaire using a visual analog scale response format from 0 to 100 (worst–best

outcome) [81]. The iHOT-33 was developed with the cooperation of the multicenter arthroscopy of the hip outcomes research network (MAHORN). It has a short version: the iHOT12 that includes 12 items instead of the original 33, designed to be more easily used in clinical settings and validated and tested for reliability. The appropriate population for this tools includes patients aged between 18 and 60 years who have a Tegner activity level of 4 or higher, meaning that they are engaged in recreational physical activities at least once a week or have an occupation involving moderately heavy labor. There were no floor or ceiling effects noted for iHOT-33 in their original paper. In the end the construct validity was demonstrated with a correlation of 0.81 to the NAHS [57].

Conditions: Femoroacetabular impingement, hip arthroscopic surgery for intra-articular hip lesion.

46.5.7 The Copenhagen Hip and Groin Outcome Score (HAGOS)

The HAGOS is a patient-reported outcome questionnaire; it consists of 37 items distributed in six subscales of pain, symptoms, physical function in activities of daily living, physical function in sports and recreation, participation in physical activities, and hip- and/or groin-related quality of life [107].

The Copenhagen hip and groin outcome score was developed in 2011, and this was the first outcome measure developed with the COSMIN checklist guidelines. The goal of this instrument is to evaluate hip and/or groin disability related to impairment (body structure and function), activity (activity limitations), and participation (participation restrictions) according to the international classification of functioning, disability, and health (ICF) in young to middle-aged physically active patients with hip and/or groin pain. The HAGOS has excellent test-retest reliability properties; this was evident from ICC ranging from 0.82 to 0.92 for all its subscales from their original paper, and it has also excellent

internal consistency properties and good content validity. Floor or ceiling effects were noted in some subscales of HAGOS as described in their original paper, while there were no floor effects for HAGOS ceiling effects that were noted in HAGOS ADL (32%) and physical activity (28%) subscales between 12 and 24 months after surgery. In the end construct validity is satisfactory between the HAGOS and the SF-36 [57, 91].

Conditions: Young to middle-aged patients with long-standing hip and/or groin pain.

46.6 Measures of Knee Function

The knee is one of the most investigated joints in Orthopaedics and Sports Medicine; therefore many outcome measures exist for clinical or research use. The most relevant are those evaluating pain, function, quality of life, and activity. However, the clinician-reported scales that have been described allow to register objective characteristics of the joint such as deformity, ROM, and stability.

The basic psychometric characteristics, strength, and weakness of the most common scales for knee function are described (Table 46.4).

46.6.1 International Knee Documentation Committee Subjective Score (Subjective IKDC)

The subjective IKDC form is an 18-item patient-reported score that examines knee symptoms, sport participation, and daily activities [53]. It was developed in 1994 and revised to its current form in 2001. Its strength is the comprehensive assessment of the patient status and, above all, responsiveness to change following surgical interventions. Its limited administrative and respondent burden makes it ideal in both clinical and research settings. Moreover, it has a good correlation with the Cincinnati knee rating system, VAS for pain, WOMAC, Lysholm, and SF-36. On the other hand, it lacks in the psychometric testing, which makes it suboptimal for the evaluation of osteoarthritic patients [21].

Conditions: Knee ligament injury and surgery, meniscal tears, cartilage lesions, knee dislocation.

46.6.2 International Knee Documentation Committee Objective Score (Objective IKDC)

The objective IKDC form is a clinician-reported score that evaluates all the aspects of knee findings. Twenty-five items, grouped in seven subgroups evaluating effusion, passive motion deficit, ligament examination, crepitus, harvest site pathology, X-ray findings, and functional evaluation with one-leg hop test, compose it. Every item is rated in a four-grade Likert scale from A to D or normal to severely abnormal, with respect to the contralateral healthy knee. The overall score is determined by the lowest value of considering only the first three subgroups (swelling, passive ROM, and ligament stability). However, all the items should be compiled even if they did not contribute to the overall score. This form is frequently used when evaluating ligamentous surgery and allows the comparison of different groups of treatment in a reliable manner. However, to increase its precision, it requires instrument-assisted evaluation of knee stability. Moreover, in the case of bilateral pathology or contralateral previous injury, the score could not be used since it implies the comparison to a healthy contralateral limb. Usually, the grade C and D are considered as failure of the treatment [1].

Conditions: Especially knee ligament injury and surgery, but also other knee conditions investigated by subjective IKDC form such as meniscal tears, cartilage lesions, knee dislocation.

46.6.3 Knee Injury and Osteoarthritis Outcome (KOOS)

The KOOS is a 42-item patient-reported scale, which includes five domains, each one scored separately: pain, symptoms, activity of daily life (ADL), sports and recreational activities, and

Table 46.4 Measures of knee function

Questionnaire	Items	Options	Administration	Range	Cutoffs	Time compilation (min)	Time calculation (min)	Internal consistency	Test-retest	SRM	MDC	MCID
Subjective IKDC ^a	18	Likert (5–10)	Patient	0 (worst) to 100 (best)	No	10	5	0.92–0.97	0.87–0.89	0.94–1.5	6.7	11.5
Objective IKDC ^b	25	Likert (4)	Clinician	A (best) to D (worst)	Yes	15	3	NR	NR	NR	NR	NR
KOOS ^a	25	Likert (5)	Patient	0 (worst) to 100 (best)	No	10	5	0.25–0.90	0.0–0.97	0.61–2.12	5–21	NR
Lysholm ^a	8	Likert (3–6)	Patient	0 (worst) to 100 (best)	Yes	5	3	0.65–0.73	0.88–0.97	0.90–1.10	8.9–10.1	NR
OKS ^a	12	Likert (5)	Patient	0 (best) to 48 (worst)	No	5	5	0.97–0.93	0.91–0.94	0.7	6.1	NR
CKRS ^b	13	Mix	Patient + clinician	0 (worst) to 100 (best)	Yes	20	10	NR	0.75–0.98	0.72–2.48	NR	NR
WOMAC ^c	24	Likert (5)	Patient	0 (best) to 20 pain, 8 stiff, 68 funct (worst)	No	10	5	0.67–0.98	0.65–0.98	0.40–2.02	10.6–30.6	14–33
WOMAC VAS ^a	24	VAS/NRS	Patient	0 (best) to 500 pain, 200 stiff, 1700 funct (worst)	No	5	3	NR	NR	NR	NR	NR
KSS ^c	7	Likert (2–25)	Patient + clinician	0 (worst) to 100 (best)	Yes	10	5	0.74–0.94	0.65–0.88	NR	NR	NR
HSS ^d	13	Mix	Patient + clinician	0 (worst) to 100 (best)	Yes	10	5	0.70	0.98–0.99	NR	NR	NR
AKPS ^e	13	Likert (3–5)	Patient	0 (worst) to 100 (best)	No	5	5	NR	0.81	NR	NR	8–10
VISA-P ^f	8	Mix	Patient	0 (worst) to 100 (best)	No	5	5	0.71–0.73	0.74	NR	NR	14

Note: Subjective IKDC (international knee documentation committee subjective score), Objective IKDC (international knee documentation committee objective score), KOOS (knee injury and osteoarthritis outcome), OKS (Oxford knee score), CKRS(Cincinnati knee rating system), WOMAC (Western Ontario and McMaster Universities index), KSS (knee society score), HSS (hospital for special surgery score), AKPS (Kujala anterior knee pain scale), VISA-P (Victorian Institute of Sport Assessment – Patella)

^aCollins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function. *Arthritis Care Res (Hoboken)*. 2011 Nov;63 Suppl 11:S208–28

^bBarber-Westin SD, Noyes FR, McCloskey JW. Rigorous statistical reliability, validity, and responsiveness testing of the Cincinnati knee rating system in 350 subjects with uninjured, injured, or anterior cruciate ligament-reconstructed knees. *Am J Sports Med*. 1999 Jul-Aug;27(4):402–16

^cHanamoto Y, Ito H, Furu M, Ishikawa M, Azukizawa M, Kuriyama S, Nakamura S, Matsuda S. Cross-cultural adaptation and validation of the Japanese version of the new Knee Society Scoring System for osteoarthritic knee with total knee arthroplasty. *J Orthop Sci*. 2015 Sep;20(5):849–53

^dNarin S, Unver B, Bakurhan S, Bozan O, Karatosun V. Cross-cultural adaptation, reliability and validity of the Turkish version of the Hospital for Special Surgery (HSS) Knee Score. *Acta Orthop Traumatol Turc*. 2014;48(3):241–8

^eCrossley KM, Bennell KL, Cowan SM, Green S. Analysis of outcome measures for persons with patellofemoral pain: which are reliable and valid? *Arch Phys Med Rehabil*. 2004 May;85(5):815–22

^fHernandez-Sanchez S, Hidalgo MD, Gomez A. Responsiveness of the VISA-P scale for patellar tendinopathy in athletes. *Br J Sports Med*. 2014 Mar;48(6):453–7

knee-related quality of life (QoL). It is a complete questionnaire, since it explores all the possible domains of a multitude of possible knee pathologies [99]. However, for this reason, acceptability and reliability could be different based on the patient's age and condition, especially on the sport subscale. It has good correlation with the SF-36 score and the WOMAC. For these reasons and its relative simplicity, the KOOS is used extensively, especially in large-volume registries. Moreover, the individual score for each subscale, rather than an aggregate score, allows for clinical interpretation of different interventions in different dimensions. On the other hand, the KOOS has not been validated for telephone and interview administration, which could limit its applicability due to the need for direct patient involvement [21].

A short version of the KOOS which includes only seven items from the ADL and sport subscales is available as the KOOS-PS (physical function short form), which is shorter, faster, and easier to administer in clinical setting [89].

Conditions: Young and middle-aged patients with post-traumatic OA (undergoing TKA), patient with chondral, ligamentous, or meniscal injuries.

46.6.4 Lysholm Score

The Lysholm score is an eight-item patient-reported scale that evaluates knee symptoms such as limp, locking, swelling, instability, pain, stair climbing, and squatting. It is one of the most commonly used clinical scores for knee evaluation, introduced in 1982 [74]. It is extremely popular and widely used in clinical and research settings. It has a limited floor and ceiling effect, making it useful for tracking improvement of interventions or deterioration over time. Moreover, it has a good correlation with the subjective IKDC, Cincinnati knee ligament score, and the WOMAC. However, its main limitation is that it is clinician derived with no patient input. There are concerns about limited reliability and a lack of definition of MCID. The Lysholm score is

graded as excellent (95–100), good (84–94), fair (65–83), or poor (<65) [21].

Conditions: Ligament injuries and surgery, particularly knee conditions with symptoms of instability, but also meniscal tears, cartilage lesions, patellofemoral pain, and knee osteoarthritis.

46.6.5 Oxford Knee Score (OKS)

The OKS is a 12-item patient-reported score developed for patients undergoing TKR [34]. However, it could be used to evaluate OA and early OA. For these reasons it has a good correlation with WOMAC, KOOS, and SF-36 scores. It is valid, reliable and responsive to change of score, which makes it useful in research settings. However, its development based on knee OA limits its use [21].

Conditions: TKR, OA, rheumatoid arthritis.

46.6.6 Cincinnati Knee Rating System (CKRS)

The CKRS, that has been proposed in 1983 and further modified along the years, is a both patient- and clinician-reported form that is composed of 13 scales assessing symptoms (pain, swelling, and giving way), perception of overall knee condition, daily life function (walking, stairs climbing, and squatting), sport function (running, jumping, and pivoting), sport activity, and occupation. The evaluation is completed with physical examination, functional testing with one-legged hoop exercises, and radiographic measurement of joint narrowing. The overall score, rated from 0 to 100, is obtained combining symptoms (20), functional activities (15), physical examination (25), stability (20), radiographic findings (10), and functional testing (10). Based on the results, it could be graded as excellent (>80), good (55–79), fair (30–54), or poor (<30). It is a comprehensive and rigorous scale, with a good reliability and high responsiveness to detect changes. However, it can be quite time-consuming. Its use

is mostly limited to sports medicine ligamentous and meniscal knee conditions [4].

Conditions: Ligamentous injuries and surgery, meniscal allograft and repair.

46.6.7 Western Ontario and McMaster Universities Index (WOMAC)

The WOMAC is a 24-item patient-reported scale that evaluates three domains, each one with a dedicated subscale: pain, stiffness, and functional activities. It is available both as five-point Likert scale and 100-mm VAS or NRS; therefore, based on the type of scoring, different ratings are obtained for the three subscales. However, the obtained values could be converted to a simple 0–100 scale. The WOMAC is one of the most common scales to evaluate patients with knee OA and is validated in numerous languages. Moreover, it has the advantage of being validated for use in person, over the telephone, or electronically. The individual scores for the three domains, rather than the aggregate value, enhance its interpretation for each domain. However, the presence of items related to uncommon tasks could result in missing data, while the lack of difficult tasks makes it not optimal for more active patients. This scale is optimal for research purpose due to its reliability and ability to detect changes especially after surgical and nonsurgical interventions for knee OA and chondral defects [6, 7, 21].

Conditions: Knee OA, cartilage lesions, and ACL injury.

46.6.8 Knee Society Score (KSS)

The KSS is a seven-item both patient- and clinician-based score that integrates subjective assessment of pain with objective features such as flexion or extension lag, ROM, alignment, and laxity. For this reason, it is limited by low reliability and inter- and intra-observer variations.

It is used mostly for TKA evaluation, and it has a good correlation with SF-36 and the OKS score. Based on the obtained values, it could be interpreted as excellent (80–100), good (70–79), fair (60–69), or poor (<60) [43, 102].

Conditions: Knee OA.

46.6.9 Hospital for Special Surgery Score (HSS)

The HSS is a 13-item scale, both patient- and clinician-reported. It evaluates pain, function, ROM, muscle strength, flexion deformity, instability and alignment. It has similar features to the KSS score and, also, could be graded as excellent (85–100), good (70–84), fair (60–69), or poor (<60). It offers a precise evaluation of knee function but lacks in general quality of life assessment. Therefore, it should be used with other scores capable of depicting patient's general condition [84].

Conditions: Knee OA.

46.6.10 Kujala Anterior Knee Pain Scale (AKPS)

The AKPS, also known as “Kujala score,” is a 13-item patient-reported scale that evaluates the subjective response to six activities, regarded as triggers for anterior knee pain such as walking, running, jumping, climbing stairs, squatting, and sitting [66]. Moreover, the evaluation is integrated with objective basic knee characteristics such as swelling, thigh atrophy, flexion contracture, and patellar abnormal movements. Therefore, it is specifically dedicated to painful conditions of the anterior knee and specifically patellofemoral pathologies. It is a simple and fast questionnaire and has a good correlation with Lysholm, KOOS, and VAS pain; however, it does not distinguish between patients with one episode of patellar dislocation and recurrent instability [27].

Conditions: Anterior knee pain conditions, patellofemoral pathologies, especially instability.

46.6.11 Victorian Institute of Sport Assessment-Patella (VISA-P)

The VISA-P score is an eight-item patient-reported questionnaire composed of a VAS and a Likert portion, which assesses pain during activity or functional tests and sport participation. It has been specifically developed for the measurement of patellar tendon-related conditions. It has a good reliability and repeatability and has a good correlation with the VAS pain. Moreover, since its MCID is available, the VISA-P represents one of the most utilized scores for the assessment of treatments for patellar tendinopathy [48].

Conditions: Patellar tendon disorders.

46.7 Measures of Foot and Ankle Function

An extremely wide range of outcome measures have been developed for the evaluation of the foot and ankle in clinical research; during a 10-year period, 139 different scales have been described, and more recently, between 2012 and 2016, as many as 89 measures have been used in literature for this anatomical region. This incredible variety might be detrimental for an evidence-based decision-making and for comparing clinical results [50, 51].

The basic psychometric characteristics, strength, and weakness of the most common scales for foot and ankle function are described (Table 46.5).

46.7.1 American Orthopaedic Foot and Ankle Society Score (AOFAS Score)

First introduced in 1994, the AOFAS score is the most used outcomes measure tools among clinicians. Four questionnaires are present for different parts of the foot: ankle/hind foot, mid-foot, hallux, and lesser toe; each one is composed of nine items divided into three domains (function, alignment, and pain) and rated on a scale from 0

to 100 [45, 64]. The AOFAS scores are not purely patient-reported since it incorporates both subjective and objective data that requires the clinical assessment. Despite its popularity the AOFAS score has limitations due to lack of validation, high inter-observer variability, and poor correlation with other generic PROMs. For these reasons the AOFAS society itself recommended the usage of more validated and standardized outcome scores [90].

Conditions: These region-specific questionnaires have been used to evaluate patients in a wide variety of foot and ankle pathologies such as arthritis, cartilage defects, soft tissue pathologies, and toe and finger deformities.

46.7.2 American Academy of Orthopaedic Surgeons: Foot and Ankle Model (AAOS-FAM)

The AAOS-FAM was released in 2004, and it is a patient-reported questionnaire composed of 25 items divided into 5 subscales: pain, function, stiffness and swelling, giving way, and shoe comfort. Each answer is measured on a scale of 1–5 or 6 and then calculated; the result is a percentage (0–100) where higher numbers represent better function. This scale is increasing in popularity among surgeons; it has good reliability and repeatability [56, 116].

Conditions: AAOS-FAM can be used to compare clinical outcomes in specific foot and ankle pathologies or surgical methods.

46.7.3 Foot Function Index (FFI)

The FFI was developed in 1991 for senior patients with foot-related pathologies; it was considered specific for foot- and ankle-related conditions secondary to rheumatoid arthritis although there is no specific item for this condition in the questionnaire. It is composed of 23 patient-reported questions that assess foot function in three domains: pain, disability, and activity limitation. It has moderate to high correlation with SF-36; this

Table 46.5 Measures of foot and ankle function

Questionnaire	Items	Options	Administration	Range	Cutoffs	Time compilation (min)	Time calculation (min)	Internal consistency	Test-retest	SRM	MDC	MCID
AOFAS ^a	9	Likert (3–4)	Patient + clinician	0 (worst) to 100 (best)	No	10	3	0.585 (0.863 for function subscale)	0.89–0.97	0.79	1.7	NR
AAOS-FAM ^b	25	Likert (5–6)	Patient	0% (worst) to 100% (best)	No	15	10	0.83–0.93	0.79–0.99	NR	NR	NR
FFI ^c	23	VAS	Patient	0 (best) to 230 (worst)	No	5	7	0.73–0.96	0.70–0.87	NR	7	NR
FAOS ^b	42	Likert (5)	Patient	0 (worst) to 100 (best)	No	10	10	0.88–0.94	0.70–0.87	NR	NR	NR
FAAM ^b	29	Mix	Patient	0% (worst) to 100% (best)	No	7	10	0.96–0.98	0.87–0.89	NR	NR	8 ADL/9 sport
FADI ^b	34	Likert (5)	Patient	0% (worst) to 100% (best)	No	7	10	0.84–0.89	NR	NR	NR	NR
ACFAS ^c	16	Likert (5)	Patient + clinician	0 (worst) to 100 (best)	No	15	10	NR	NR	NR	NR	NR
FHSQ ^b	13	Likert (3–5)	Patient	0 (worst) to 100 (best)	No	6	10	0.89–0.95	0.74–0.92	NR	NR	7–14
ROFPAQ ^b	39	Likert (5)	Patient	1 (best) to 5 (worst)	No	10	6	0.81–0.89	0.82–0.93	NR	NR	NR
QOL ^b	5	Likert (5)	Patient	0 (worst) to 20 (best)	No	3	1	0.85–0.91	NR	NR	NR	NR
OMAS ^d	9	Likert (3–5)	Patient	0 (worst) to 100 (best)	No	3	3	0.76	0.95	NR	NR	NR
VISA-A ^e	8	Mix	Patient	0 (worst) to 100 (best)	No	5	7	0.73	0.79	NR	NR	NR

Note: AOFAS (American orthopaedic foot and ankle society score), AAOS-FAM (American Academy of Orthopaedic: foot and ankle), FFI (foot function index), FAOS (foot and ankle outcomes score), FAAM (foot and ankle ability measure), FADI (foot and ankle disability index), ACFAS (American College of Foot and Ankle Surgeons), FHSQ (foot health status questionnaire), ROFPAQ (Rowan foot pain assessment), QOL (sport ankle QOL), OMAS (Olerud-Molander ankle score), VISA-A (Victorian Institute of sports assessment-Achilles)

^aDe Boer AS, Meuffels DE, et al. Validation of the American Orthopaedic Foot and Ankle Society Ankle-Hindfoot Scale Dutch language version in patients with hindfoot fractures. *BMJ Open*. 2017;7(11):e018314

^bMartin RL, Irrgang JJ. A survey of self-reported outcome instruments for the foot and ankle. *J Orthop Sports Phys Ther*. 2007;37(2):72–84

^cCook JJ, Cook EA, et al. Validation of the American College of Foot and Ankle Surgeons Scoring Scales. *J Foot Ankle Surg*. 2011;50(4):420–9

^dNilsson et al. The Swedish version of OMAS is a reliable and valid outcome measure for patients with ankle fractures. *BMC Musculoskeletal Disorders*. 2013;14:109

^eIversen, J. V., Bartels, et al. Danish VISA-A questionnaire. *Scand J Med Sci Sports*. 2016; 26: 1423–1427

suggests that FFI may be a good measure of both health status and patients' outcomes [13, 14, 104].

Conditions: Generally used in older patients, rheumatoid patients, orthotics outcomes, poor reliability in professional athlete due to the reported ceiling effect.

46.7.4 Foot and Ankle Outcomes Score (FAOS)

The FAOS, released in 2001, is a 42-item patient-reported outcomes measure that consists in five subscales (pain, symptoms, ADL, sport, and ankle-related quality of life). Each subscale is graded separately and scored in a 0–100 value. The FAOS demonstrated good reliability and validity, but the length of this survey can create significant burden for the patient [41, 98].

Conditions: It has been validated for a variety of foot and ankle pathologies such as adult flat-foot deformity, hallux valgus, hallux rigidus.

46.7.5 Foot and Ankle Ability Measure (FAAM)

The FAAM was developed in 2005; it is region-specific and composed of 29 patient-reported items divided in activity of daily living (ADL) and sport subscales. A recent study demonstrated that the FFI and FAAM are highly correlated for foot and ankle trauma patients [40, 77].

Conditions: It is valid for a range of foot and ankle conditions as well as for chronic ankle instability and diabetes mellitus-related conditions.

46.7.6 Foot and Ankle Disability Index (FADI)

The FADI was first released in 1999; it is the former version of FAAM and includes four more items for pain assessment and one item for the ability to sleep (34 total items). FADI and FAAM are appropriate to evaluate functional disabilities in athletes with chronic ankle instability [76].

Conditions: Sport-related foot and ankle pathologies and trauma evaluation.

46.7.7 American College of Foot and Ankle Surgeons (ACFAS) Universal Evaluation Scoring Scales

The American College of Foot and Ankle Surgeons developed these anatomically based scoring scales in 2005 as clinical instruments to evaluate objective and subjective parameters before and after surgery [106]. Four modules exist for the first metatarsal-phalangeal joint and first ray, the forefoot (excluded first ray), the rear foot, and the ankle; each questionnaire is completed by both patient and clinician and includes subjective (pain, appearance, and functional capacity) and objective (radiographic and functional) parameters, for a total of 100 points. This instrument has been validated and presents good reliability and sensitivity to change [24].

Conditions: Foot and ankle musculoskeletal-related pathologies requiring surgical intervention.

46.7.8 Foot Health Status Questionnaire (FHSQ)

The FHSQ was developed for individuals undergoing surgical treatment for common foot conditions. It consists of four subscales with a total of 13 items representing the following four domains: pain (four items), function (four items), footwear (three items), and general foot health (two items). Scores from each subscale range from 0 to 100, with a higher score representing better outcomes [8].

Conditions: Foot- and ankle-related disorders including those affecting skin and nail.

46.7.9 Rowan Foot Pain Assessment (ROFPAQ)

The ROFPAQ was developed as a disease-specific instrument for chronic foot pain. It contains

39 items in the following four subscales for pain assessment: sensory (16 items), affective (ten items), cognitive (ten items), and comprehension (three items). Each subscale is scored independently from 1 through 5, and the item responses are merged together to produce a subscale score ranging from 1 to 5, with a higher score representing more pain [100].

Conditions: Chronic foot and ankle pain.

46.7.10 Sport Ankle QOL (Quality of Life)

The sport ankle rating system quality of life measure was developed as a region-specific measure including self-reported and clinician-completed outcome measures. The QOL measure, the clinical rating score, and a single numeric evaluation are the three outcome measures that could be used together or independently. The QOL is a self-reported questionnaire designed to assess an athlete's quality of life after an ankle injury; it contains five items that evaluate symptoms, work and school activities, recreation and sports activities, activities of daily living, and lifestyle.

The clinical rating score is composed of 11 items both patient and clinician based; finally, with the numeric VAS evaluation, the patient is asked to score his ankle function from 0 to 100 [113].

Conditions: Ankle injuries and specifically ankle sprains.

46.7.11 Olerud-Molander Ankle Score (OMAS)

The OMAS is a disease-specific outcomes measure developed for patients with ankle fractures and has been frequently used to evaluate this group of subjects; furthermore, it has been reported to be a valid item for recording short-term changes after an acute ankle ligament injury. OMAS is a self-administered patient questionnaire; the scale ranges from 0 points (totally impaired function) to 100 points (completely unimpaired function) and is based on nine differ-

ent domains: pain, stiffness, swelling, stair climbing, running, jumping, squatting, supports, and work/activity level [87, 88].

Conditions: Ankle fracture, ligament ankle injury.

46.7.12 Victorian Institute of Sports Assessment-Achilles (VISA-A)

The VISA-A is a disease-specific instrument designed to evaluate the clinical severity for patients with chronic Achilles tendinopathy. It is an easily self-administered questionnaire that evaluates symptoms and their effect on physical activity. The questionnaire contains eight questions, covering three necessary domains: pain, functional status, and activity. The first six questions use a visual analog scale so that the patient may report the magnitude of a continuum of subjective symptoms; the final two questions used a categorical rating scale. The final results range from 0 to 100, with asymptomatic persons expected to score 100 points [95].

Conditions: Chronic Achilles tendinopathy.

46.8 Measures of Activity Level

Making patients capable of an unlimited physical activity is the main focus of clinicians; for this reason several scores have been created to assess outcomes in terms of return to sport/activity (RTS). While considering these instruments, a factor to be outlined is that athletes are different from the general population since they have higher level of physical function and perceived health, often they do not perceive symptoms during the daily activities, and common outcome measures may not detect problems that only result from high-intensity training and competition.

The basic psychometric characteristics, strength, and weakness of the most common scales for activity level are described below (Table 46.6).

Table 46.6 Measures of activity level

Questionnaire	Items	Options	Administration	Range	Cutoffs	Time compilation (min)	Time calculation (min)	Internal consistency	Test-retest	SRM	MDC	MCID
TEGNER ^a	1	Likert (11)	Patient	0 (worst) to 10 (best)	No	1	1	NR	0.8	1.0	1	1
UCLA ^b	1	Likert (10)	Patient	1 (worst) to 10 (best)	Yes	2	2	NR	NR	NR	1	1
ARS ^c	4	Likert (5)	Patient	0 (worst) to 16 (best)	No	1	1	8.87	0.81	NR	NR	NR
AAS ^d	1	Likert (11)	Patient	0 (worst) to 10 (best)	No	1	1	1.0	NR	NR	1	1
LEFS ^e	20	Likert (5)	Patient	0 (worst) to 80 (best)	No	5	3	0.96	0.86	NR	9	9
SAS ^f	7	Likert (4–5)	Patient	0 (worst) to 20(best)/A–D	No	1	1	NR	0.92	NR	NR	NR
HAS ^g	44	Likert (6)	Patient	0 (worst) to 220 (best)	No	120	30	NR	0.75	NR	NR	NR
OSTRC ^h	4 (each joint)	Likert (4–5)	Patient	0 (best) to 100 (worst)	No	>5 (variable)	3–7	0.86–0.91	NR	NR	NR	NR
SQUASH ⁱ	11	Open	Patient	1 min. to 9 max. (per item) no upper limit	No	3–5	10	NR	NR	NR	NR	NR
IPAQ-SF ^j	7	Open	Patient	0 to No upper limit	No	3–5	10	<0.8	NR	NR	NR	NR
HAP ^k	94	Likert (3)	Patient	1 (worst) to 94 (best)	No	10	5	NR	0.84 MAS; 0.79 AAS	NR	6.5 MAS; 8.4 AAS	NR

Note: ARS (activity rating scale), AAS (ankle activity score), LEFS (lower extremity functional scale), SAS (shoulder activity scale), HAS (Heidelberg sport activity score), OSTRC (Oslo Sports Trauma Research Centre overuse injury questionnaire), SQUASH (short questionnaire to assess health-enhancing physical activity), IPAQ-SH (international physical activity questionnaire – short form), HAP (human activity profile)

^aBriggs KK, Lysholm J, et al. The reliability, validity, and responsiveness of the Lysholm score and Tegner activity scale for anterior cruciate ligament injuries of the knee: 25 years later. *Am J Sports Med.* 2009;37(5):890–7

^bTerwee CBI, Bouwmeester W, et al. Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Osteoarthritis Cartilage.* 2011;19(6):620–33

^cHosseini Negahban, Neda Mostafaei, Soheil Mansour Sohani, Masood Mazaheri, Shahin Goharpey, Mahyar Salavati, Shabla Zahednejad, Zohreh Meshkati & Ali Montazeri (2011) Reliability and validity of the Tegner and Marx activity rating scales in Iranian patients with anterior cruciate ligament injury, *Disability and Rehabilitation*, 33:22–23, 2305–2310

^dTamás Halasi, MD, Ákos Kynsburg, MD, et al. Development of a New Activity Score for the Evaluation of Ankle Instability. *The American Journal of Sports Medicine.* 2004;32:899–908

^eAlcock GK, Stratford PW. Validation of the Lower Extremity Function Scale on athletic subjects with ankle sprains. *Physiother Can.* 2002;54:233–240

^fBrophy RH, Beauvais RL, et al. Measurement of shoulder activity level. *Clin Orthop Relat Res.* 2005;439:101–8

^gSeeger J, Weinmann S, et al. The Heidelberg Sports Activity Score - A New Instrument to Evaluate Sports Activity. *The Open Orthopaedics Journal.* 2013;7:25–32

^hClarsen B, Myklebust G, et al. Development and validation of a new method for the registration of overuse injuries in sports injury epidemiology: the Oslo Sports Trauma Research Centre (OSTRC) Overuse Injury Questionnaire. *Br J Sports Med.* 2013;47:495–502

ⁱTerwee CB, Bouwmeester W, et al. Instruments to assess physical activity in patients with osteoarthritis of the hip or knee: a systematic review of measurement properties. *Osteoarthritis Cartilage.* 2011;19(6):620–33

46.8.1 Tegner Activity Score

First described in 1985 [105] for the prospective evaluation of the knee ligaments injuries, the Tegner activity scale provides an arbitrary ranking based on the level of sport and leisure time activities and competition at which the patient is currently participating. It is a simple scale in which the subject indicates his/her current activity ranging from 0 (no physical activity/disabled) to 10 (participation in competitive soccer or pivoting sports). It was created as a complement to the Lysholm score; but its use has also extended into other joints, including the hip and ankle [42, 78].

Conditions: Tegner score was developed and is mostly used for knee ligamentous injuries and reconstructions.

46.8.2 University of California at Los Angeles (UCLA) Activity Rating Scale

The UCLA activity rating scale is a simple scale ranging from 1 (no activity) to 10 (participation in impact sports); it was developed in 1998 to assess physical activity after joint replacements. Like the Tegner score, the patient is asked to rate his/her own most appropriate activity level. Four activity subgroups were defined: scores between 0 and 4 (low activity), 4.1 and 6 (moderately low activity), 6.1 and 8 (moderately high activity), and 8.1 and 10 (high activity) [114].

Conditions: The UCLA is mostly used and validated for hip and knee osteoarthritis and evaluation of joint replacement.

46.8.3 Activity Rating Scale (ARS) or Marx Scale

The ARS/Marx questionnaire quantifies the frequency of activities that challenge the dynamic stability of the knee; it consists in four questions about how frequently the patients perform activity such as running, cutting, decelerating, and pivoting. Each question is scored from 0 (<1

time/month) to 4 (>4 time/month), and the total score range is 0–16. ARS is based on the idea of measuring specific components of function/movement (that apply universally to the lower limb) to allow more accurate comparison among patients. This scale can be completed in a very short time frame [78].

Conditions: Sport activity involving complex articular motions of the knee and lower limb.

46.8.4 Ankle Activity Score (AAS)

The AAS is a joint-specific score that was published in 2004; it was based on the Tegner score. It contains 53 sports, three working activities, and four general activities; the patient is asked to select his/her most appropriate sport/activity and to indicate a level of participation (top level, lower competitive level, recreational level). The result, as with the Tegner score, is represented by a single number from 0 to 10 [42].

Conditions: Ankle injuries.

46.8.5 Lower Extremity Functional Scale (LEFS)

The LEFS was developed to be a broad region-specific measure for individuals with musculoskeletal disorders of the hip, knee, ankle, or foot. It consists of 20 items that specifically cover the domains of activity and participation. The scale uses a Likert response format, with a higher score representing a higher level of ability [10].

Conditions: LEFS have been validated for several pathologies of the lower limb; moreover, it has been translated in different languages.

46.8.6 Shoulder Activity Scale (SAS)

The Brophy-Marx SAS was developed in 2005 as an easy instrument to evaluate the patient's overall shoulder activity level that could be generalized across different sports and completed in less than 1 min. It is composed of two parts: the first five items describe five common activities of the

shoulder and the relative frequency, during the patient's previous year, for each item is scored from 0 to 4 (never or less than once a month, once a month, once a week, more than once a week, or daily). The total numerical activity score ranges from a minimum of 0 points to a maximum of 20 points. In the second part of the score, the patients are asked if they participate in contact sports and sports that involve repetitive overhead throwing. The answers of these two questions range from A (no) to D (yes, at professional level). SAS has shown good reliability and validity [11, 12].

Conditions: This score has been developed on patients with rotator cuff tears.

46.8.7 Heidelberg Sport Activity Score (HSAS)

The HAS was published in 2013; this validated instrument divides sport activities into 11 categories: walking, swimming, cycling, running, cross-country skiing, alpine skiing, golfing, dancing, racket sports, ball sports, and miscellaneous. For each of these activities, the patient is asked to grade between 0 and 5 about frequency, duration, level of importance, and impairment from the affected joint. For each activity, a core from 0 to 20 is calculated with a formula: $(\text{frequency} + \text{duration}) \times (1 + \text{impairment}/10 + \text{importance}/10)$. The scores are then added to obtain a final score between 0 and 220. HAS has proven high validity, sensitivity, reliability, and sensitivity. It can be used for elite-level athletes and athletes who perform different sports and is valid for different joints; nevertheless, its disadvantage is an extremely long time for compilation (120 min) [103].

Conditions: Evaluation of activity after trauma or surgery; it can be used in elite-level athletes.

46.8.8 Oslo Sports Trauma Research Center Overuse Injury Questionnaire (OSTRC)

The intention of the OSTRC was to create a questionnaire that could be applied to overuse

injury problems in any area of the body. The instrument is designed in four items for each affected joint; the final "severity score" ranges between 0 and 100 (25 point for item) for each overuse problem. In studies with multiple anatomical areas of interest, the four questions are repeated for each area. This questionnaire uses the term "problem" rather than "injury" since there is greater variation in interpretation of the term "injury" [20].

Conditions: OSTRC is mainly used for the evaluation of overuse problems in sports injury epidemiology.

46.8.9 Short Questionnaire to Assess Health-Enhancing Physical Activity (SQUASH)

The SQUASH was not designed to measure energy expenditure but to give an indication of the habitual activity level. It consists of 11 questions on commuting activities, leisure time and sports activities, household activities, and activities at work and school. The total activity score is calculated by taking the sum of the activity scores for separate questions [112].

Conditions: SQUASH is a short physical activity questionnaire with the general purpose to assess habitual physical activity.

46.8.10 International Physical Activity Questionnaire-Short Form (IPAQ-SF)

The IPAQ-SF consists in seven questions about the frequency and durations of participation in strenuous, moderate, and walking activities in addition to the time spent sitting during the past week. The final score is expressed in metabolic equivalents (METs) which represent the oxygen consumption of an individual sitting for 1 min (3.5 mL/kg/min) [26, 29].

Conditions: IPAQ has been validated, and it presents reasonable measurement properties for monitoring population levels of physical activity in diverse settings.

46.8.11 Human Activity Profile (HAP)

The HAP is a 94-item self-report measure of energy expenditure or physical fitness; it was developed as an outcome measure for medical rehabilitation for people with a wide spectrum of physical disorders. It consists of a list of activities for which patients should indicate if they are currently able to perform the activity, have stopped performing the activity, or have never performed the activity. Each of the selected activities has an estimate energy requirement between approximately 1 and 10 METs. Two scores are calculated: the maximum activity score (MAS) and the adjusted activity score (AAS) [38].

Conditions: Epidemiologic and population studies as well as rehabilitation medicine.

Fact Box 46.2

Athletes population is in many ways different from general one; they have much higher level of physical functioning and health status. For this reason, they do not perceive symptoms during the daily activities and choosing an adequate outcome measure is compulsory.

46.9 Measures of Global and Mental Health

Generic measures of health-related quality of life are frequently used to evaluate the impact of treatments and clinical results and to monitor population health. Often these scales are composed of various independent domains/dimensions that together represent the notion of health-related quality of life. The items are weighted to indicate the relative importance attributed to them by the respondents and then aggregated into a single number reflecting the quality or value of a health state. To obtain such values, several instruments have been developed.

The basic psychometric characteristics, strength, and weakness of the most common scales for global and mental health are described (Table 46.7).

46.9.1 36-Item Short-Form Health Survey (SF-36) and Short-Form 12 (SF-12)

The SF-36 is a general health measure, introduced in 1992 that includes 36 items addressing eight domains of overall health status: physical functioning (PF), bodily pain (BP), role limitations due to physical health problems (RP), role limitations due to personal or emotional problems (RE), general mental health (MH), social functioning (SF), energy/fatigue or vitality (VIT), and general health perceptions (GH). Although this scale has been validated for orthopedic use, experts recommend pairing the SF-36 with an orthopedic-specific measure since it is a general health scale, and it could be difficult to isolate orthopedic outcomes from other unrelated health conditions.

The SF-12 is a shortened version of the SF-36, developed in 1996, with the aim of reducing redundancies and time burden on the patient. It shortens the survey to 12 items and reports 2 scores in physical and mental domains. It has been validated for orthopedic patients. The SF-12 was included as a recommended PRO measure for “general quality of life” by the AAOS.

Both SF-36 and SF-12 are great questionnaires for outcomes assessment in research; both need to be administered in conjunction with orthopedic-specific measures [15, 52, 110, 111, 115].

46.9.2 EuroQol-5 Domains-3 Likert (EQ-5D-3L)

The EQ-5D health status and quality-of-life measure is composed of five items (mobility, self-care, usual activity, pain/discomfort, and anxiety/depression), with three possible response levels (no problems, some/moderate problems, extreme problems). The EQ-5D index is

Table 46.7 General and mental health measures

Questionnaire	Items	Options	Administration	Range	Cutoffs	Time compilation (min)	Time calculation (min)	Internal consistency	Test-retest	SRM	MDC	MCID
SF-36 ^a	36	Likert (2–6)	Patient	0 (worst) to 100 (best)	No	7	13	≥0.70 (0.90 for physical function)	<0.70	1.04	5	NR
SF-12 ^a	12	Likert (2–6)	Patient	0 (worst) to 100 (best)	No	3	7	≥0.82 PCS; ≥0.75 MCS	0.89 PCS; 0.79 MCS	<0.2	NR	NR
EQ-5D-3L ^b	6	Mix	Patient	−0.594 (worst) to 1 (best)/0–100	No	5	10	NR	NR	0.7/0.04	NR	NR
AQoL ^c	12	Likert (4)	Patient	−0.04 (worst) to 1.0 (best)	No	5	7	NR	NR	NR	NR	NR
NPH ^a	45	Dichotomous	Patient	0 (best) to 100 (worst)	No	5–10	10–15	0.83 (0.33 for mobility)	0.77–0.85	NR	NR	NR
PROMIS 10 ^b	10	Mix	Patient	t-score distribution	No	5	2	NR	NR	0.72	NR	NR

Note: SF-36 (36-Item Short Form Health Survey), SF-12 (12-Item Short Form Health Survey), EQ-5D-3L (EuroQoL-5 Domains-3 Likert), AQoL (Assessment of Quality of Life), NPH (Nottingham Health Profile), PROMIS (Patient-Reported Outcomes Measurement Information System 10 Global health)

^aBusija, L., Pausenberger, E., et al. (2011), Adult measures of general health and health-related quality of life: Medical Outcomes Study Short Form 36-Item (SF-36) and Short Form 12-Item (SF-12) Health Surveys, Nottingham Health Profile (NHP), Sickness Impact Profile (SIP), Medical Outcomes Study Short Form 6D (SF-6D), Health Utilities Index Mark 3 (HUI3), Quality of Well-Being Scale (QWB), and Assessment of Quality of Life (AQOL). *Arthritis Care Res*, 63: S383-S412

^bOak SR, Strnad GJ, et al. Responsiveness Comparison of the EQ-5D, PROMIS Global Health, and VR-12 Questionnaires in Knee Arthroscopy. *Orthop J Sports Med*. 2016 Dec 17;4(12):2325967116674714

^cKathryn Whitfield, Rachelle Buchbinder, et al. Parsimonious and efficient assessment of health-related quality of life in osteoarthritis research: validation of the Assessment of Quality of Life (AQoL) instrument. *Health and Quality of Life Outcomes* 2006, 4:19

calculated from the five dimensions, ranging from -0.594 (worst) to 1.0 . Moreover, to the EQ-5D index, the EQ-5D includes a VAS for rating of overall health status from 0 (worst imaginable health) to 100 (best imaginable health). A common criticism of this measure is the lack of sensitivity to change since only three levels of responses are available within each construct. With the aim of addressing this issue, a version of the measure with five responses has been developed, called the EQ-5D-5L [47].

46.9.3 Assessment of Quality of Life (AQoL)

The AQoL is a 12-item instrument which loads onto four dimensions: independent living, social relationships, physical senses, and psychological well-being. These subscales are weighted between 0.0 (death) and 1.0 (full health). With its emphasis upon psychosocial dimensions of health, it offers significant advantages for evaluation studies where these dimensions are important [93].

46.9.4 Nottingham Health Profile (NPH)

The NHP questionnaire is a self-administered questionnaire. It was developed in English and consists of two parts: Part 1 contains 38 “yes/no” questions covering six dimensions: pain, physical mobility, emotional reactions, energy, social isolation, and sleep. Part 2 has seven “yes/no” questions concerning problems of daily activities. It has been shown to be internally consistent, valid, reproducible, and sensitive [55].

46.9.5 Patient-Reported Outcomes Measurement Information System 10 Global Health (PROMIS-10 Global Health)

The PROMIS was established in 2004 with funding from the National Institutes of Health; this

initiative develops and evaluates standard measures for key patient-reported health indicators and symptoms. PROMIS measures are standardized, allowing for assessment of many patient-reported outcome domains such as pain, fatigue, emotional distress, physical functioning, and social role participation.

Computerized adaptive testing (CAT) software has been implemented; this allows tailoring the PRO assessment to the individual patient by selecting the most informative set of questions based on responses to previous questions [17].

46.10 Measures of Pain

Pain is a complex and subjective experience and that implies several measurement challenges. It is important for the clinicians to utilize sensitive and accurate pain outcome measures although currently we rely mainly on self-report measures. The cutoff value for clinical significance of pain reduction must be determined on the minimal amount of change being important to patients. A reduction of 10–20% of pain can be considered clinically significant [67].

The basic psychometric characteristics, strength, and weakness of the most common scales for pain assessment are described (Table 46.8).

46.10.1 Visual Analog Scale for Pain (VAS for Pain)

The VAS for pain was introduced in 1976; it is a widely recognized and simple instrument that allows the patient to score his own pain level on a straight 100-mm line with zero indicating “no pain” and 100 “worst imaginable pain.” Its usefulness in orthopedic surgery has been recognized, and VAS has proven high validity and responsiveness; on the other hand, its low specificity has been shown with a 1.1 cm decrease corresponding to the minimal clinical important difference for pain. The patient acceptable symptomatic state is considered with a value less than 3 cm [37, 46, 59, 115].

Table 46.8 Pain measures

Questionnaire	Items	Options	Administration	Range	Cutoffs	Time compilation (min)	Time calculation (min)	Internal consistency	Test-retest	SRM	MDC	MCID
VAS ^a	1	VAS	Patient	0 mm (best) to 100 mm (worst)	Yes	<1	<1	NR	0.71–0.94	NR	1 mm	1.1 cm
NRS ^a	1	Likert (11)	Patient	0 (best) to 10 (worst)	No	<1	<1	NR	0.96	NR	1	2
VRS ^a	1	Likert (5)	Patient	1 (best) to 5 (worst)	No	<1	<1	NR	NR	NR	1	1
FPS-R ^b	1	Likert (6)	Patient	0 (best) to 10 (worst)	No	<1	<1	NR	NR	NR	2	2
SF-MPQ ^a	17	Mix	Patient	0–45/0–5/0–10 (best) to (worst)	No	2–5	1–2	0.73–0.89	0.79–0.93	>0.80	5.2, 4.5, 2.8, 1.4, 1.4 cm	>5

Note: VAS (visual analog scale for pain), NRS (numerical rating scale for pain), VRS (verbal rating scale for pain), FPS-R (faces pain scale revised), SF-MPQ (short-form McGill pain questionnaire)

^aHawker, G. A., Mian, S., et al. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short-Form McGill Pain Questionnaire (SF-MPQ), Chronic Pain Grade Scale (CPGS), Short Form-36 Bodily Pain Scale (SF-36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care Res*, 2011;63: S240-S252

^bFerreira-Valente MA, Pais-Ribeiro JL, et al. Validity of four pain intensity rating scales. *Pain*. 2011;152(10):2399–404

46.10.2 Numerical Rating Scale for Pain (NRS for Pain)

The NRS is an 11-point scale consisting of integers from 0 through 10: 0 representing “no pain” and 10 representing “worst imaginable pain.” Respondents select the single number that best represents their pain intensity. It is considered to be more comprehensive compared to the VAS for; however, it may capture the complex nature of the pain experience [46].

46.10.3 Verbal Rating Scale for Pain (VRS for Pain)

The VRS is a single domain five-point scale consisting of a list of sentences (no pain, mild pain, moderate pain, intense pain, maximum pain) describing increasing levels of pain severity. Respondents select the single phrase that best characterizes their pain intensity [46].

46.10.4 Faces Pain Scale-Revised (FPS-R)

The FPS-R is a six-point scale represented by six different faces showing increasing severity of pain. Patients are asked to select the facial expression that best resembles his or her pain intensity, from the left-most face (“no pain”) to the right-most face (“very much pain”). The FPS-R was originally developed for pediatric patients, but its simplicity makes it a reliable instrument for individuals with cognitive and communication impairments as well [9, 49, 101].

46.10.5 Short-Form McGill Pain Questionnaire (SF-MPQ)

The SF-MPQ is a multidimensional measure, with extensive clinical research use. Patients rate their pain in sensory terms (e.g., sharp or stabbing) and affective terms (e.g., sickening or fearful), with 15 total descriptors. Each item is rated on a four-point scale that ranges from none to

severe. The SF-MPQ also has a single VAS item for pain intensity and a VRS for rating the overall pain experience. It is used particularly to measure the sensory and affective aspects of pain and pain intensity in adults with chronic pathologies [46, 79].

Fact Box 46.3

Since “pain” perception is maybe the most relevant outcome in clinical practice, research protocols should include sensitive and accurate pain measures. Currently we rely mainly on self-report measures where a reduction of 10–20% of pain can be considered as the minimal amount of change being important to patients.

46.11 Measures of Sport-Related Psychological Aspects

It has been demonstrated that while most athletes reach a normal physical function, less than half of them return to the same level of sport activity. Possibly, psychological factors are involved in the rehabilitation processes and in the athlete’s self-perception of recovery. The following section describes some of the common scales for sport activity assessment, energy expenditure, and psychological factors after injuries.

The basic psychometric characteristics, strength, and weakness of the most common scales for sport-related psychological aspects are described (Table 46.9).

46.11.1 Injury-Psychological Readiness to Return to Sport Scale (I-PRRS)

The I-PRRS is an easy to use tool developed to measure the athlete’s psychological readiness to return to sport after injury. It is a six-item scale, each item is scored from 0 (no confidence) to 100 (maximum confidence) with intervals of 10. The scores from the six items are summed and divided

Table 46.9 Psychological scores and measures

Questionnaire	Items	Options	Administration	Range	Cutoffs	Time compilation (min)	Time calculation (min)	Internal consistency	Test-retest	SRM	MDC	MCID
IPRRS ^a	6	Likert (11)	Patient	0 (worst) to 60 (best)	Yes	8–10	5–8	0.63	0.97	NR	NR	NR
RIAI ^b	28	Likert (4)	Patient	0 (best) to 45 (worst)	No	8–15	8	0.96–0.98	NR	NR	NR	NR
TSK ^c	11	Likert (4)	Patient	11 (best) to 44 (worst)	No	4–8	5	0.7–0.9	>0.7	NR	NR	NR

Note: IPRRS (injury-psychological readiness to return to sport scale), RIAI (re-injury anxiety inventory), TSK (Tampa scale for kinesiophobia)

^aNaghdi S, Nakhostin Ansari N, et al. Cross-cultural adaptation and validation of the Injury-Psychological Readiness to Return to Sport scale to Persian language. *Physiother Theory Pract.* 2016;32(7):528–35

^bWalker, Natalie et al. A preliminary development of the Re-Injury Anxiety Inventory (RIAI). *Physical Therapy in Sport*, Volume 11, Issue 1, 23–29

^cRoelofs J, van Breukelen G, et al. Norming of the Tampa Scale for Kinesiophobia across pain diagnoses and various countries. *Pain.* 2011;152:1090–1095

by 10 to calculate the total score. The range of scores is between 0 and a maximum score of 60. A score of 60 indicates high confidence to return to sport; 40, moderate confidence; and 20, low confidence [39].

Conditions: Evaluation of psychological readiness to return to sport among athletes.

46.11.2 Re-injury Anxiety Inventory (RIAI)

The RIAI is a 28-item score designed to assess the athlete's fear of experiencing a re-injury. It is composed of the rehabilitation anxiety (RIA-R, 13 items) and the reentry to competition anxiety (RIA-RE, 15 items). The instrument is based on a four-point (0–3) Likert-response type; the final score ranges from 0 (complete absence of anxiety) to 45 (extreme anxiety) [109].

Conditions: RIAI can be used in studies aiming to evaluate athlete's psychological readiness for RTS.

46.11.3 Tampa Scale for Kinesiophobia (TSK)

The TSK is a self-reported measure developed to assess "fear of movement-related pain" in patients with musculoskeletal disorders. The original test, developed in English 1, has been translated into ten languages. The TSK-11 is the most widely used; it contains 11 items from the original 17-item questionnaire. Each item is scored on a four-point Likert scale, ranging from 1 "strongly disagree" to 4 "strongly agree"; total scores vary between 11 and 44, with higher scores indicating higher levels of fear of movement-related pain [96].

Conditions: The TSK-11 is a reliable and valid measurement tool that provides therapists valuable information on activity avoidance and pathological somatic focus in patients with musculoskeletal pain.

Clinical Vignette

An innovative and minimally invasive surgical technique for complex osteochondral knee lesions was developed in your institution. From the first outpatient follow-ups, you realize that the subjects treated with this new technique seem to be very happy about their health status and knee function. Finally, you are charged to design a study protocol to compare the result of the new technique with the classic one. Beside an accurate imaging of the bone and cartilage and maybe a biochemical characterization of the fluids, which clinical outcomes measures can be included in our protocol? Which ones are more indicated to detect an effective improvement in patient conditions?

First one or more knee specific measures should be chosen. In this pathology, the KOOS score has demonstrated good psychometric properties, and the WOMAC score has an excellent reliability and ability to detect changes.

Secondly, we want to assess the patient's perceived pain level and health status. For this purpose, the SF-MPQ score for pain is accurate and easy to complete; moreover it includes a VAS for general pain assessment, while the SF-12 score has a short-time compilation and will give a precise overview of patient's general health.

Finally, since most of our patients used to be quite active before injury, our protocol should include at least one activity level measure. We choose the Tegner score, which was specifically designed for knee injuries and is extremely intuitive in its compilation, and the LEFS score which can be used for several lower limb pathologies and shows good reliability.

Take-Home Message

- The development, testing, and implementation of tools to aid in the measurement of phenomena in medicine are central to clinical practice and clinical research; therefore PROMs are a key component to orthopedics research and may also be so for clinical practice in orthopedic surgery.
- Health status measurement instruments must possess adequate measurement properties, and it is fundamental to remember that a complete picture of a condition or a treatment effect on a patient could be provided only with the combination of a “disease-specific measure” and a “generic measure.”

References

1. Aichroth PM, Cannon WD Jr. International Knee Documentation Committee: knee ligament injury and reconstruction 209 evaluation. Knee surgery current practice. New York: Raven Press; 1992. p. 759–60.
2. Amstutz HC, Sew Hoy AL, Clarke IC. UCLA anatomic total shoulder arthroplasty. *Clin Orthop*. 1981;155:7–20.
3. Angst F, Schwyzler HK, et al. Measures of adult shoulder function. *Arthritis Care Res*. 2011;63(Suppl 11):S174–88.
4. Barber-Westin SD, Noyes FR, et al. Rigorous statistical reliability, validity, and responsiveness testing of the Cincinnati knee rating system in 350 subjects with uninjured, injured, or anterior cruciate ligament-reconstructed knees. *Am J Sports Med*. 1999;27(4):402–16.
5. Beaton DE, Wright JG, Katz JN, The Upper Extremity Collaborative Group. Development of the QuickDASH: comparison of three item-reduction approaches. *J Bone Joint Surg Am*. 2005;87:1038–46.
6. Bellamy N. WOMAC Osteoarthritis Index user guide. London: University of Western Ontario; 1995. p. 79.
7. Bellamy N. WOMAC Osteoarthritis Index user guide. Version V. Brisbane: CONROD, The University of Queensland; 2002.
8. Bennett PJ, Patterson C, et al. Development and validation of a questionnaire designed to measure foot-health status. *J Am Podiatr Med Assoc*. 1998;88:419–28.
9. Bieri D, Reeve R, et al. The Faces Pain Scale for the self-assessment of the severity of pain experienced by children: development, initial validation and preliminary investigation for ratio scale properties. *Pain*. 1990;41:139–50.
10. Binkley JM, Stratford PW, et al. The Lower Extremity Functional Scale (LEFS): scale development, measurement properties, and clinical application. *Phys Ther*. 1999;79:371–83.
11. Brophy RH, Beauvais RL, et al. Measurement of shoulder activity level. *Clin Orthop Relat Res*. 2005;439:101–8.
12. Brophy RH, Levy B, et al. Shoulder activity level varies by diagnosis. *Knee Surg Sports Traumatol Arthrosc*. 2009;17:1516–21.
13. Budiman-Mak E, Conrad K, et al. Theoretical model and Rasch analysis to develop a revised Foot Function Index. *Foot Ankle Int*. 2006;27(7):519–27.
14. Budiman-Mak E, Conrad KJ, et al. The Foot Function Index: a measure of foot pain and disability. *J Clin Epidemiol*. 1991;44(6):561–70.
15. Busija L, Osborne RH, et al. Magnitude and meaningfulness of change in SF-36 scores in four types of orthopaedic surgery. *Health Qual Life Outcomes*. 2008;6:55.
16. Celik D. Psychometric properties of the Mayo Elbow Performance Score. *Rheumatol Int*. 2015;35(6):1015–20.
17. Cella D, Riley W, et al. Initial Adult Health Item Banks and First Wave Testing of the Patient-Reported Outcomes Measurement Information System (PROMISTM) network: 2005–2008. *J Clin Epidemiol*. 2010;63(11):1179–94.
18. Christensen CP, Althausen PL, Mittleman MA, et al. The nonarthritic hip score: reliable and validated. *Clin Orthop Relat Res*. 2003;406:75–83.
19. Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. *J Hand Surg Am*. 1998;23(4):575–87.
20. Clarsen B, Myklebust G, et al. Development and validation of a new method for the registration of overuse injuries in sports injury epidemiology: the Oslo Sports Trauma Research Centre (OSTRC) Overuse Injury Questionnaire. *Br J Sports Med*. 2013;47:495–502.
21. Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS), and Tegner Activity Score (TAS).

- Arthritis Care Res (Hoboken). 2011;63(Suppl 11):S208–28. <https://doi.org/10.1002/acr.20632>.
22. Constant CR, Gerber C, Emery RJ, Sojbjerg JO, Gohlke F, Boileau P. A S186 Angst et al review of the Constant score: modifications and guidelines for its use. *J Shoulder Elb Surg*. 2008;17:355–61.
 23. Constant CR, Murley AH. A clinical method of functional assessment of the shoulder. *Clin Orthop Relat Res*. 1987;214:160–4. 56.
 24. Cook JJ, Cook EA, et al. Validation of the American College of Foot and Ankle Surgeons Scoring Scales. *J Foot Ankle Surg*. 2011;50(4):420–9.
 25. Cooney WP, Bussey R, Dobyns JH, Linscheid RL. Difficult wrist fractures. Perilunate fracture-dislocations of the wrist. *Clin Orthop Relat Res*. 1987;(214):136–47.
 26. Craig CL, Marshall AL, et al. International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc*. 2003;35(8):1381–95.
 27. Crossley KM, Bennell KL, et al. Analysis of outcome measures for persons with patellofemoral pain: which are reliable and valid? *Arch Phys Med Rehabil*. 2004;85(5):815–22.
 28. Dacombe PJ, Amirfeyz R, Davis T. Patient-reported outcome measures for hand and wrist trauma: is there sufficient evidence of reliability, validity, and responsiveness? *Hand (New York NY)*. 2016;11(1):11–21.
 29. Daughton DM, Fix AJ, et al. Maximum oxygen consumption and the ADAPT quality-of life scale. *Arch Phys Med Rehabil*. 1982;63:620–2.
 30. Dawson J, Doll H, Boller I, et al. The development and validation of a patient-reported questionnaire to assess outcomes of elbow surgery. *J Bone Joint Surg Br*. 2008;90:466–73.
 31. Dawson J, Fitzpatrick R, Carr A, Murray D. Questionnaire on the perceptions of patients about total hip replacement. *J Bone Joint Surg Br*. 1996;78:185–90.
 32. Dawson J, Fitzpatrick R, Carr A. Questionnaire on the perceptions of patients about shoulder surgery. *J Bone Joint Surg Br*. 1996;78:593–600.
 33. Dawson J, Fitzpatrick R, Carr A. The assessment of shoulder instability. The development and validation of a questionnaire. *J Bone Joint Surg Br*. 1999;81:420–6.
 34. Dawson J, Fitzpatrick R, Murray D, Carr A. Questionnaire on the perceptions of patients about total knee replacement. *J Bone Joint Surg Br*. 1998;80:63–9.
 35. Dawson J, Rogers K, Fitzpatrick R, Carr A. The Oxford Shoulder Score revisited. *Arch Orthop Trauma Surg*. 2009;129:119–23.
 36. Dreiser R, Maheu E, Guillou GB, Caspard H, Grouin JM. Validation of an algofunctional index for osteoarthritis of the hand. *Rev Rhum Engl Ed*. 1995;62:43S–53S.
 37. Ferreira-Valente MA, Pais-Ribeiro JL, et al. Validity of four pain intensity rating scales. *Pain*. 2011;152(10):2399–404.
 38. Fix AJ, Daughton DM. Human activity profile professional manual. Odessa: Psychological Assessment Resources, Inc.; 1988.
 39. Glazer DD. Development and preliminary validation of the Injury-Psychological Readiness to Return to Sport (I-PRRS) Scale. *J Athl Train*. 2009;44(2):185–9.
 40. Goldstein CL, Schemitsch E, et al. Comparison of different outcome instruments following foot and ankle trauma. *Foot Ankle Int*. 2010;31(12):1075–80.
 41. Golightly YM, Devellis RF, et al. Psychometric properties of the foot and ankle outcome score in a community-based study of adults with and without osteoarthritis. *Arthritis Care Res*. 2014;66(3):395–403.
 42. Halasi T, Kynsburg A, et al. Development of a new activity score for the evaluation of ankle instability. *Am J Sports Med*. 2004;32:899–908.
 43. Hamamoto Y, Ito H, et al. Cross-cultural adaptation and validation of the Japanese version of the new Knee Society Scoring System for osteoarthritic knee with total knee arthroplasty. *J Orthop Sci*. 2015;20(5):849–53.
 44. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. *J Bone Joint Surg Am*. 1969;51:737–55.
 45. Hasenstein T, Greene T, et al. A 5-year review of clinical outcome measures published in the Journal of the American Podiatric Medical Association and the Journal of Foot and Ankle Surgery. *J Foot Ankle Surg*. 2017;56(3):519–21.
 46. Hawker GA, Mian S, et al. Measures of adult pain: Visual Analog Scale for Pain (VAS Pain), Numeric Rating Scale for Pain (NRS Pain), McGill Pain Questionnaire (MPQ), Short Form McGill Pain Questionnaire (SF MPQ), Chronic Pain Grade Scale (CPGS), Short Form 36 Bodily Pain Scale (SF 36 BPS), and Measure of Intermittent and Constant Osteoarthritis Pain (ICOAP). *Arthritis Care Res*. 2011;63:S240–52.
 47. Herdman M, Gudex C, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Qual Life Res*. 2011;20(10):1727–36.
 48. Hernandez-Sanchez S, Hidalgo MD, et al. Responsiveness of the VISA-P scale for patellar tendinopathy in athletes. *Br J Sports Med*. 2014;48(6):453–7.
 49. Hicks CL, von Baeyer CL, et al. The Faces Pain Scale – Revised: toward a common metric in pediatric pain measurement. *Pain*. 2001;93:173–83.
 50. Hunt KJ, Hurwit D. Use of patient-reported outcome measures in foot and ankle research. *J Bone Joint Surg Am*. 2013;95(16):e118(1–9).
 51. Hunt KJ, Lakey E. Patient-reported outcomes in foot and ankle surgery. *Orthop Clin North Am*. 2018;49(2):277–89.

52. American Academy of Orthopaedic Surgeons. Instruments for collection of orthopaedic quality data; 2016.
53. Irrgang JJ, Anderson AF, Boland AL, Harner CD, Kurosaka M, Neyret P, et al. Development and validation of the International Knee Documentation Committee subjective knee form. *Am J Sports Med.* 2001;29:600–13.
54. ISAKOS Scientific Committee, Audigé L, Ayeni OR, Bhandari M, Boyle BW, Briggs KK, Chan K, Chaney-Barclay K, Do HT FM, Fu FH, Goldhahn J, Goldhahn S, Hidaka C, Hoang-Kim A, Karlsson J, Krych AJ, RF LP, Levy BA, Lubowitz JH, Lyman S, Ma Y, Marx RG, Mohtadi N, Marcheggiani Muccioli GM, Nakamura N, Nguyen J, Poehling GG, Roberts LE, Rosenberg N, Shea KP, Sohani ZN, Soudry M, Voineskos S, Zaffagnini S, International Society of Arthroscopy, Knee Surgery and Orthopaedic Sports Medicine. A practical guide to research: design, execution, and publication. *Arthroscopy.* 2011;27(4 Suppl):S1–112.
55. Jenkinson C, Fitzpatrick R, et al. The Nottingham Health Profile: an analysis of its sensitivity in differentiating illness groups. *Soc Sci Med.* 1988;27:1411–4.
56. Johanson NA, Liang MH, et al. American Academy of Orthopaedic Surgeons lower limb outcomes assessment instruments. Reliability, validity, and sensitivity to change. *J Bone Joint Surg Am.* 2004;86-a(5):902–9.
57. Kemp JL, Collins NJ, Roos EM, Crossley KM. Psychometric properties of patient-reported outcome measures for hip arthroscopic surgery. *Am J Sports Med.* 2013;41:2065–73.
58. Kennedy CA, Beaton DE, Solway S, McConnell S, Bombardier C. The DASH outcome measure user's manual. 3rd ed. Toronto: Institute for Work & Health; 2011.
59. Kersten P, White PJ, et al. Is the pain visual analogue scale linear and responsive to change? An exploration using Rasch analysis. *PLoS One.* 2014;9(6):e99485.
60. Kirkley A, Griffin S, Alvarez C. The development and evaluation of a disease-specific quality of life measurement tool for rotator cuff disease: the Western Ontario Rotator Cuff Index (WORC). *Clin J Sport Med.* 2003;13:84–92.
61. Kirkley A, Griffin S, et al. Scoring systems for the functional assessment of the shoulder. *Arthroscopy.* 2003;19(10):1109–20.
62. Kirkley A, Griffin S, McLintock H, Ng L. The development and evaluation of a disease-specific quality of life measurement tool for shoulder instability: the Western Ontario Shoulder Instability Index (WOSI). *Am J Sports Med.* 1998;26:764–72.
63. Kirkley A, Werstine R, Ratjek A, Griffin S. Prospective randomized clinical trial comparing the effectiveness of immediate arthroscopic stabilization versus immobilization and rehabilitation in first traumatic anterior dislocations of the shoulder: long-term evaluation. *Arthroscopy.* 2005;21:55–63.
64. Kitaoka HB, Alexander I, et al. Clinical rating systems for the ankle-hindfoot, midfoot, hallux, and lesser toes. *Foot Ankle Int.* 1994;15(7):349–53.
65. Klassbo M, Larsson E, Mannevik E. Hip Disability and Osteoarthritis Outcome Score: an extension of the Western Ontario and McMaster Universities Osteoarthritis Index. *Scand J Rheumatol.* 2003;32:46–51.
66. Kujala UM, Jaakola LH, Koskinen SK, Taimela S, Hurme M, Nelimarkka O. Scoring of patellofemoral disorders. *Arthroscopy.* 1993;9:159–63.
67. Lee JS, Hobden E, et al. Clinically important change in the visual analogue scale after adequate pain control. *Acad Emerg Med.* 2003;10(10):1128–30.
68. Lequesne M. Indices of severity and disease activity for osteoarthritis. *Semin Arthritis Rheum.* 1991;20(Suppl 2):48–54.
69. Lequesne MG, Mery C, Samson M, Gerard P. Indexes of severity for osteoarthritis of the hip and knee: validation—value in comparison with other assessment tests. *Scand J Rheumatol Suppl.* 1987;65:85–9.
70. Lequesne MG. The algofunctional indices for hip and knee osteoarthritis. *J Rheumatol.* 1997;24:779–81.
71. Lippitt SB, Harryman DT, Matsen FA. A practical tool for evaluation of function: the Simple Shoulder Test. In: Matsen FA, Fu FH, Hawkins RJ, editors. *The shoulder: a balance of mobility and stability.* Rosemont: American Academy of Orthopedic Surgeons; 1993. p. 545–59.
72. Lo IKY, Griffin S, Kirkley A. The development and evaluation of a disease-specific quality of life measurement tool for osteoarthritis of the shoulder: The Western Ontario Osteoarthritis of the Shoulder Index (WOOS). *Osteoarthritis Cartilage.* 2001;9:771–8.
73. Longo G, Franceschi F, Loppini M, Maffulli N, Denaro V. Rating systems for evaluation of the elbow. *Br Med Bull.* 2008;87(1):131–61.
74. Lysholm J, Gillquist J. Evaluation of knee ligament surgery results with special emphasis on use of a scoring scale. *Am J Sports Med.* 1982;10:150–4.
75. Macdermid J. Update: the Patient-rated Forearm Evaluation Questionnaire is now the Patient-rated Tennis Elbow Evaluation. *J Hand Ther.* 2005;18:407–10.
76. Martin R, Burdett R, et al. Development of the Foot and Ankle Disability Index (FADI). *J Orthop Sports Phys Ther.* 1999;29(1):A32–3.
77. Martin RL, Irrgang JJ, et al. Evidence of validity for the Foot and Ankle Ability Measure (FAAM). *Foot Ankle Int.* 2005;26(11):968–83.
78. Marx RG, Stump TJ, et al. Development and evaluation of an activity rating scale for disorders of the knee. *Am J Sports Med.* 2001;29(2):213–8.
79. Melzack R. The short-form McGill Pain Questionnaire. *Pain.* 1987;30:191–7.
80. Michener LA, McClure PW, Sennett BJ. American Shoulder and Elbow Surgeons Standardized Shoulder Assessment Form, patient self-report

- section: reliability, validity, and responsiveness. *J Shoulder Elb Surg.* 2002;11(6):587–94.
81. Mohtadi NGH, Griffin DR, Pedersen ME, et al. The development and validation of a self-administered quality-of-life outcome measure for young, active patients with symptomatic hip disease: the International Hip Outcome Tool (iHOT-33). *Arthroscopy.* 2012;28(5):595–610.e1.
 82. Morrey BF, An KN. Functional evaluation of the elbow. In: Morrey BF, editor. *The elbow and its disorders.* Philadelphia: WB Saunders; 1993.
 83. Murray DW, Fitzpatrick R, Rogers K, Pandit H, Beard DJ, Carr AJ, et al. The use of the Oxford Hip and Knee Scores. *J Bone Joint Surg Br.* 2007;89:1010–4.
 84. Narin S, Unver B, et al. Cross-cultural adaptation, reliability and validity of the Turkish version of the Hospital for Special Surgery (HSS) Knee Score. *Acta Orthop Traumatol Turc.* 2014;48(3):241–8.
 85. Nilsson AK, Lohmander LS, Klassbo M, Roos EM. Hip Disability and Osteoarthritis Outcome Score (HOOS): validity and responsiveness in total hip replacement. *BMC Musculoskelet Disord.* 2003;4:10.
 86. Nilsson A, Bremander A. Measures of hip function and symptoms: Harris Hip Score (HHS), Hip Disability and Osteoarthritis Outcome Score (HOOS), Oxford Hip Score (OHS), Lequesne Index of Severity for Osteoarthritis of the Hip (LISOH), and American Academy of Orthopedic Surgeons (AAOS) Hip and Knee Questionnaire. *Arthritis Care Res.* 2011;63:S200–7.
 87. Nilsson, et al. The Swedish version of OMAS is a reliable and valid outcome measure for patients with ankle fractures. *BMC Musculoskelet Disord.* 2013;14:109.
 88. Olerud C, Molander H. A scoring scale for symptom evaluation after ankle fracture. *Arch Orthop Trauma Surg.* 1984;103:190–4.
 89. Perruccio AV, Stefan Lohmander L, Canizares M, Tennant A, Hawker GA, Conaghan PG, et al. The development of a short measure of physical function for knee OA KOOS-Physical Function Shortform (KOOS-PS): an OARSI/OMERACT initiative. *Osteoarthritis Cartil.* 2008;16:542–50.
 90. Pinski E, Daniels TR. AOFAS position statement regarding the future of the AOFAS clinical rating systems. *Foot Ankle Int.* 2011;32(9):841–2.
 91. Ramisetty N, Kwon Y, et al. Patient-reported outcome measures for hip preservation surgery: a systematic review of the literature. *J Hip Preserv Surg.* 2015;2(1):15–27.
 92. Richards RR, An KN, Bigliani LU, Friedman RJ, Gartsman GM, Gristina AG, et al. A standardized method for the assessment of shoulder function. *J Shoulder Elb Surg.* 1994;3:347–52.
 93. Richardson J, Sinha K, et al. Modeling the utility of health states with the Assessment of Quality of Life (AQoL) 8D instrument: overview and utility scoring algorithm. Research paper, vol. 63. Melbourne: Centre for Health Economics, Monash University; 2011.
 94. Roach KE, Budiman-Mak E, Songsiridej N, Lertratanakul Y. Development of a Shoulder Pain and Disability Index. *Arthritis Care Res.* 1991;4:143–9.
 95. Robinson JM, Cook JL, et al. The VISA-A questionnaire: a valid and reliable index of the clinical severity of Achilles tendinopathy. *Br J Sports Med.* 2001;35(5):335–41.
 96. Roelofs J, van Breukelen G, et al. Norming of the Tampa Scale for Kinesiophobia across pain diagnoses and various countries. *Pain.* 2011;152:1090–5.
 97. Rompe JD, Overend TJ, et al. Validation of the Patient-rated Tennis Elbow Evaluation Questionnaire. *J Hand Ther.* 2007;20(1):3–10.
 98. Roos EM, Brandsson S, et al. Validation of the foot and ankle outcome score for ankle ligament reconstruction. *Foot Ankle Int.* 2001;22(10):788–94.
 99. Roos EM, Roos HP, Lohmander LS, Ekdahl C, Beynon BD. Knee Injury and Osteoarthritis Outcome Score (KOOS): development of a self-administered outcome measure. *J Orthop Sports Phys Ther.* 1998;28:88–96.
 100. Rowan K. The development and validation of a multi-dimensional measure of chronic foot pain: the Rowan Foot Pain Assessment Questionnaire (ROFPAQ). *Foot Ankle Int.* 2001;22:795–809.
 101. Scott J, Huskisson EC. Graphic representation of pain. *Pain.* 1976;2(2):175–84.
 102. Scuderi GR, Bourne RB, Noble PC, Benjamin JB, Lonner JH, Scott WN. The New Knee Society Knee Scoring System. *Clin Orthop Relat Res.* 2012;470(1):3–19.
 103. Seeger J, Weinmann S, et al. The Heidelberg Sports Activity Score - a new instrument to evaluate sports activity. *Open Orthopaed J.* 2013;7:25–32.
 104. SooHoo NF, Vyas R, et al. Responsiveness of the foot function index, AOFAS clinical rating systems, and SF-36 after foot and ankle surgery. *Foot Ankle Int.* 2006;27(11):930–4.
 105. Tegner Y, Lysholm J. Rating systems in the evaluation of knee ligament injuries. *Clin Orthop Relat Res.* 1985;(198):43–9.
 106. Thomas JL, Christensen JC, et al. ACFAS Scoring Scale user guide. *J Foot Ankle Surg.* 2005;44(5):316–35.
 107. Thorborg K, Holmich P, Christensen A, Petersen F, Roos EM. The Copenhagen Hip and Groin Outcome Score (HAGOS): development and validation according to the COSMIN checklist. *Br J Sports Med.* 2011;45:478–91.
 108. van der Linde JA, van Kampen DA, et al. The Oxford Shoulder Instability Score; validation in Dutch and first-time assessment of its smallest detectable change. *J Orthop Surg Res.* 2015;10:146.
 109. Walker N, et al. A preliminary development of the Re-Injury Anxiety Inventory (RIAI). *Phys Ther Sport.* 2010;11(1):23–9.

110. Ware J Jr, Kosinski M, et al. A 12-item short-form health survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996;34(3):220–33.
111. Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30(6):473–83.
112. Wendel-Vos GCW, Schuit AJ, et al. Reproducibility and relative validity of the short questionnaire to assess health-enhancing physical activity. *J Clin Epidemiol*. 2003;56:1163e9.
113. Williams GN, Molloy JM, et al. Evaluation of the Sports Ankle Rating System in young, athletic individuals with acute lateral ankle sprains. *Foot Ankle Int*. 2003;24:274–82.
114. Zahiri CA, Schmalzried TP, et al. Assessing activity in joint replacement patients. *J Arthroplast*. 1998;13(8):890–5.
115. Zampelis V, Ornstein E, et al. A simple visual analog scale for pain is as responsive as the WOMAC, the SF-36, and the EQ-5D in measuring outcomes of revision hip arthroplasty. *Acta Orthop*. 2014;85(2):128–32.
116. Zwiers R, Weel H, et al. Large variation in use of patient-reported outcome measures: a survey of 188 foot and ankle surgeons. *Foot Ankle Surg*. 2017. pii: S1268-7731(17)30050-4.



A Practical Guide to Writing (and Understanding) a Scientific Paper: Meta-Analyses

47

Alberto Grassi, Riccardo Compagnoni,
Kristian Samuelsson, Pietro Randelli, Corrado Bait,
and Stefano Zaffagnini

47.1 Introduction

In the era of evidence-based medicine, clinicians are continuously facing the massive production of clinical studies, often with discordant evidence. To facilitate the spread of knowledge, narrative reviews, systematic reviews and meta-analyses represent indispensable tools for summarising the evidence related to a specific topic.

In a *narrative review*, an “expert” summarises the important aspects relating to a topic. It is assumed that this expert will be objective in presenting the pertinent information. Since a specific research question and a systematic search of the

evidence are lacking, bias (usually unintentional) is often a problem [19].

Conversely, a *systematic review* represents a higher and unbiased level of evidence, since it assimilates information about a topic or question with more rigour, sophistication and transparency. A systematic review is a formal process for gathering and evaluating literature to answer a specific question, beginning with the posing of the question, the definition of the inclusion and exclusion criteria of trials and extracting the necessary data from each one. Moreover, methodological quality evaluation is usually performed [19].

A *meta-analysis* represents a further step, as it statistically combines the data from clinical

A. Grassi (✉)

Dipartimento Scienze Biomediche e Neuromotorie
DIBINEM, Università di Bologna, Bologna, Italy

Ila Clinica Ortopedica e Traumatologica, IRCCS
Istituto Ortopedico Rizzoli, Bologna, Italy

SIGASCOT Arthroscopy Committee, Florence, Italy
e-mail: alberto.grassi3@studio.unibo.it

R. Compagnoni

Società Italiana del Ginocchio Artroscopia Sport
Cartilagine Tecnologie Ortopediche (SIGASCOT)
Arthroscopy Committee, Florence, Italy

1st Department, Azienda Socio Sanitaria Territoriale
Centro Specialistico Ortopedico Traumatologico
Gaetano Pini-CTO, Milan, Italy

K. Samuelsson

Department of Orthopaedics, Institute of Clinical
Sciences, The Sahlgrenska Academy, University of
Gothenburg, Gothenburg, Sweden

Department of Orthopaedics, Sahlgrenska University
Hospital, Mölndal, Sweden

P. Randelli

1st Department, Azienda Socio Sanitaria Territoriale
Centro Specialistico Ortopedico Traumatologico
Gaetano Pini-CTO, Milan, Italy

C. Bait

Società Italiana del Ginocchio Artroscopia Sport
Cartilagine Tecnologie Ortopediche (SIGASCOT)
Arthroscopy Committee, Florence, Italy

Istituto Clinico Villa Aprica, Como, Italy

S. Zaffagnini

Ila Clinica Ortopedica e Traumatologica, IRCCS
Istituto Ortopedico Rizzoli, Bologna, Italy

Laboratori di Biomeccanica e Innovazione
Tecnologica, Istituto Ortopedico Rizzoli,
Bologna, Italy

trials, often randomised controlled trials (RCT), obtained from a systematic review of the literature. The reason for combining the data is an attempt to increase the ability to see a difference between two groups, reducing the chance of type-II errors (missing the existence of a true difference), and to increase the precision of the estimated effect by increasing sample size. Meta-analyses are therefore a powerful tool for accumulating and summarising knowledge; however, their use is also controversial, as there are several critical conditions and methodological considerations that could produce misleading conclusions. For this reason, a meta-analysis requires personal judgement and expertise and the implementation of procedures and quality standards to reduce the risk of bias that may influence the results [11, 30] (Fact Box 47.1).

The aim of this review is to provide the reader with the basic knowledge to write and understand a meta-analysis, describing all the methodological steps in its preparation and providing a useful reference for a detailed, in-depth understanding of this complex and controversial field.

This guide has been developed with four different aims:

Fact Box 47.1

- **Narrative review:** when an “expert” summarises the important aspects relating to a topic without a systematic presentation of the evidences. It is usually source of unintentional bias.
- **Systematic review:** formal process for gathering and evaluating literature to answer a specific question, beginning with the posing of the question, the definition of the inclusion and exclusion criteria of trials and extracting the necessary data from each one.
- **Meta-analysis:** process to statistically combine the data from clinical trials, often randomised controlled trials (RCT), obtained from a systematic review of the literature.

1. To provide a guide to prepare a meta-analysis for those clinicians approaching the meta-analysis study design
2. To provide a summary and a source of references for those clinicians already familiar with the production of meta-analyses
3. To help the reader understand a meta-analysis and interpret its results
4. To help the journal reviewer identify the critical aspects of meta-analyses

For a more detailed guide, consulting the *Cochrane Handbook for Systematic Reviews of Interventions* [15] and the “Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA)” [25] is warmly recommended.

47.2 How a Meta-Analysis Works: The Concept of Effect Size

To summarise, a classical meta-analysis of RCTs provides a single, overall measurement of the treatment effect, enhancing the clinical interpretation of findings across several studies [17]. This technique can prove especially useful when there are several similar clinical trials with or without consistent outcomes or when there are small- to medium-sized trials with inconclusive results. The statistical calculation behind the results of a meta-analysis is based on the concept of “effect size”. Effect size is defined as a quantitative measure of the magnitude of a phenomenon, and in statistics, it is a parameter that reflects the magnitude and direction of the treatment effect for each study. For example, if all the studies in the meta-analysis measure a continuous outcome, such as the Lysholm score after single- or double-bundle ACL reconstruction, the mean difference between the two groups can be used as the effect size and therefore to express the effects of the treatment. The overall effect size derived from the meta-analysis is calculated by combining the effect sizes of the included studies.

The magnitude and direction of the effect size are integrated with the size confidence intervals (CI). CIs are reported as a probability (e.g. 95% confidence interval, 95% CI), providing a range

with an upper and lower boundary that indicates the precision of the treatment estimates of the effect size of the included studies. A wider CI, which may be caused by a small sample size or by imprecision in the measurement (e.g. wide standard deviations in the considered outcome), indicates less precise estimates and could therefore question the application of the results in clinical practice.

47.3 How to Start Your Meta-Analysis Properly

47.3.1 Define a Study Question

The first and most important step in preparing a meta-analysis is to identify an appropriate question to address [26]. According to the *Journal of Bone and Joint Surgery* guidelines (Fact Box 47.2) [36], a meta-analysis should address a question that has not been considered in the previous 5 years, unless the most recent literature has changed dramatically. Furthermore, the question at the centre of a meta-analysis should not have already been answered satisfactorily by the results of multiple well-conducted RCTs. Finally, the question should be addressed to the evaluation of the effects of two different treatments for the same clinical condition, to allow the inclusion of only randomised or quasi-randomised clinical trials. Typical topics of meta-analyses in orthopaedic surgery could be a comparison of clinical results and the failure rate after single- or double-bundle and hamstring or bone-patellar tendon-bone anterior cruciate ligament (ACL) reconstruction, the re-rupture rate and complications after the conservative or surgical treatment of Achilles tendon ruptures, the re-dislocation rate after a brace or repair after primary patellar dislocation, clinical outcomes and alignment after patient-specific instrumentation (PSI) or conventional total knee arthroplasty (TKA), functionality and pain after hyaluronic acid or platelet-rich plasma (PRP) for knee osteoarthritis, to mention just a few.

An unconventional use of a meta-analysis could be employed to pool together the data from single-arm case series studies. Continuous mea-

Fact Box 47.2: Journal of Bone and Joint Surgery (JBJS) Requirements to Perform a Meta-analysis [36]

- JBJS will not accept meta-analyses or systematic reviews on the same topic published within 5 years unless the authors can demonstrate that the literature has dramatically changed.
- Meta-analyses or systematic reviews will not be accepted if the same (or largely the same) papers are used to arrive at similar conclusions.
- For meta-analyses in which the authors use statistical methods to combine and summarise results, only summaries of randomised trials will be accepted.
- Only studies with sufficient homogeneity of inclusion and exclusion criteria will be considered appropriate for meta-analysis.
- Authors should familiarise themselves with reporting checklists such as the PRISMA to improve the quality of reporting.

surements such as the Lysholm score, or dichotomous variables, such as return to sport [9], failures, or good vs. poor outcomes [10, 28], could also be pooled, to obtain a mean value for a wider population composed of the sum of the populations of single studies. In certain circumstance, with wide and non-heterogeneous populations, statistical calculations could be used to compare the outcomes for different groups [7]; however, the results should be interpreted with extreme caution, since the data derive from single-arm case series and not from RCTs.

47.3.2 Perform an Appropriate Literature Search

The next practical step is to obtain the largest number of studies relevant to the study question [20]. The literature search should usually be performed by two independent investigators using at least

three databases, as there is an incomplete overlap of the results from single databases. The three bibliographic databases generally regarded as being the most important sources to search for the reports of trials are CENTRAL, PubMed and EMBASE (Excerpta Medica Database). The Cochrane Central Register of Controlled Trials (CENTRAL) serves as the most comprehensive source of reports of controlled trials. CENTRAL is published as part of *The Cochrane Library* and its access is free of charge. Medical Literature Analysis and Retrieval System Online (MEDLINE) currently contains over 16 million references to journal articles from the 1950s and onwards. PubMed provides access to a free version of MEDLINE that also includes up-to-date citations not yet indexed for MEDLINE. Lastly, EMBASE currently contains over 12 million records from 1974 and onwards. EMBASE.com is Elsevier's own version of EMBASE which, in addition to the 12 million EMBASE records from 1974 and onwards, also includes over seven million unique records from MEDLINE from 1966 to the present day. Access to EMBASE is only available by subscription.

In addition to the previous three main databases, other sources should be searched, such as national and regional databases, tables of contents of relevant journals (such as the *Journal of Bone and Joint Surgery*, *Arthroscopy*, the *American Journal of Sports Medicine*, *Knee Surgery Sports Traumatology Arthroscopy*, *Clinical Orthopaedics and Related Research* for the orthopaedic field), grey literature of unpublished trials and a manual search or the reference list of the papers included in the meta-analysis. As clinical trials with positive results have more chance of being published, there could be a publication bias. For this reason, many journals now require every published RCT to be registered at the clinicaltrials.gov website before their execution, to promote the tracking of each RCT despite its positive or negative results (Fact Box 47.3). This source should also be searched to minimise the possibility of missing relevant results.

In terms of the search strategy, the choice of the key terms should aim to produce the most extensive search possible. However, it is necessary to strike a balance between striving for

Fact Box 47.3: What clinicaltrials.gov is?

ClinicalTrials.gov was created because of the Food and Drug Administration Modernization Act of 1997. It is a Web-based resource that provides patients, their family members, healthcare professionals, researchers and the public with easy access to information on publicly and privately supported clinical studies on a wide range of diseases and conditions. Each ClinicalTrials.gov record presents summary information about a study protocol including disease, intervention, study design and contact information for the study locations and, for some records, also includes information on the results of the study.

comprehensiveness and maintaining relevance, as increasing the comprehensiveness of a search will reduce its precision and retrieve more non-relevant articles. Usually, three sets of terms, developed for the healthcare condition, intervention(s) and study design, should be combined using the Boolean “AND” operator. To be as comprehensive as possible, it is necessary to include a wide range of free-text terms for each of the selected concepts (e.g. “injury”, “rupture” or “lesion” related to the ACL or Achilles tendon), combined using the Boolean “OR” operator. The function of “truncation” (e.g. Menisc* for Meniscus or Meniscal) could be a strategy for maximising the results. As this is one of the most important phases of the development of a meta-analysis, the help of a librarian specialising in database searches could be useful.

47.4 How to Obtain the Appropriate Data

47.4.1 Define Precise Inclusion and Exclusion Criteria

The most common reference manager software enables the pooling together of all the results obtained from the systematic search in the

selected databases and the discarding of all the duplicates. To prepare a precise flow chart of study selection, all the numbers of studies found, removed and excluded should be noted and justified.

The crucial step in the phase of study selection is a clear and precise definition of inclusion and exclusion criteria [26]. Their goal is to create a homogeneous study population for the meta-analysis. The rationale for their choice should be stated, as it may not be apparent to the reader. Inclusion criteria may be based on study design, sample size and characteristics of the subject, type of treatment and follow-up. Examples of exclusion criteria include studies not published in English or as full-length manuscripts, drop-out rate (usually >20%), doctoral theses and studies published in non-peer-reviewed journals. Moreover, the outcomes that should be analysed and the way they are presented could be a matter of inclusion and exclusion criteria (e.g. radiographic evaluation of osteoarthritis according only to the Kellgren-Lawrence scale). This search and data extraction (presented below) should be performed by two independent investigators; the results should therefore be compared and any disagreement should be resolved by a third independent investigator. Usually, the first screening of all the results is based on title and abstract evaluation. The full text of the potentially eligible studies should then be obtained and carefully evaluated.

47.4.2 Extract All the Relevant Information

Two or more authors of a meta-analysis should abstract information from studies independently. “Data” is defined as any information about (or deriving from) a study, including details of methods, participants, setting, context, interventions, outcomes, results, publications and investigators [13]. Data should only be collected from separate sets of patients, and the authors should be careful to avoid studies that publish the same subjects or overlapping groups of subjects that appeared in different studies in duplicate publications. In this

Table 47.1 Summary of the main items that should be extracted from each study included in a meta-analysis

Main items considered in data collection for a meta-analysis	
Methods	Study design and duration, sequence generation, allocation sequence concealment, blinding, other concerns regarding bias, inclusion and exclusion criteria
Participants	Total number, setting, diagnostic criteria, age, gender, co-morbidities
Interventions	Number of intervention groups, specific interventions, intervention details
Outcomes	Outcomes and time points, definition, unit, scale
Results	Number of participants in each intervention group and for each outcome, summary data for each intervention group, eventual subgroup analysis
Miscellaneous	Funding source, key conclusion from the study authors

case, the wider population or most recent report can be chosen. The main items considered in data collection are presented in Table 47.1:

If all the relevant information cannot be obtained from the full text, the study authors could be contacted to request the missing desired information. This step could be crucial if several parameters are missing in the outcome report, since these data are fundamental for the statistical calculation of the meta-analysis. For dichotomous outcomes, only the number of patients that experience the outcome and the total number of patients are needed. Moreover, categorical values (e.g. Kellgren-Lawrence scale for knee osteoarthritis) should be analysed as dichotomous variables, after defining a clinically meaningful cut-off value to determine two groups. On the other hand, for continuous data, the mean values and especially the SDs are needed as well. Since the SD is mandatory for effect-size calculation, if it is not reported in the original study, there are several artefacts to approximate it from the known information, such as standard error (SE), the range of values or the *p*-value [14] (Table 47.2). The last option is to impute the SD, borrowing the SD from a similar study included in the meta-analysis,

Table 47.2 Methods to derive the standard deviation from the data usually provided in a scientific paper

Obtain standard deviation from the available data
Obtaining standard deviation from the standard error
Standard deviation: Standard error $\times \sqrt{\text{sample size}}$
Obtaining standard deviation from the ranges
Standard deviation: (Upper range – lower range)/4
Obtaining standard deviation from the <i>p</i> -value
<i>Step 1: from p-value to t-value</i>
Excel formula: =tinv(<i>P</i> -value,degrees of freedom)
Degrees of freedom is The number of patients in an experimental group + control group – 2
<i>Step 2: from t-value to standard error</i>
Standard error: (mean of group 1 – means of group 2)/ <i>t</i> -value
<i>Step 3: from standard error to standard deviation</i>
Standard deviation: Standard error / $\sqrt{(1/\text{patient of group one} + 1/\text{group two})}$

using the highest or a “reasonably high” SD, or using the average SD. All imputation techniques involve making assumptions about unknown statistics introducing a source of bias, and it is best to avoid using them wherever possible. If most of the studies in a meta-analysis have missing standard deviations, these values should not be imputed, and the meta-analysis cannot be performed. A narrative presentation of the results is then preferable. On the other hand, when SDs are derived, a sensitivity analysis (see further section) is recommended to establish whether the derived imprecision could affect the result of the meta-analysis and the effect of a treatment.

47.5 How to Analyse the Obtained Data

47.5.1 Choose Appropriate Statistical Software

The tabulation of data and the calculation of simple medians, means and SDs are possible with Microsoft Excel or the equivalent. However, to perform an appropriate statistical analysis of a meta-analysis, dedicated software is necessary. One of the most frequently used is Review Manager (RevMan, Copenhagen: The Nordic Cochrane Centre, The Cochrane Collaboration, 2014), the official software of the Cochrane

Community, which is available free of charge and which has been designed to facilitate preparation of protocols and full reviews, including text, characteristics of studies, comparison tables and study data. Moreover, it can perform meta-analysis of the data entered and present the results graphically. Another valid free alternative is OpenMetaAnalyst, which is a simple open-source software with an Excel-like interface for performing meta-analyses of binary, continuous or diagnostic data, using a variety fixed- and random-effect methods, including Bayesian and maximum likelihood analysis. It also enables to perform meta-regression analysis, meta-analysis of proportions and continuous variables from single-arm studies and cumulative, leave-one-out or subgroup analyses. Another simple software is MedCalc (MedCalc Software, Ostend, Belgium), which offers also the possibility to perform a multitude of statistical tests and analysis, with a variety of graphic representations. However, its options for meta-analysis purposes are limited, and moreover it is only available freely as a sample.

47.5.2 Define Correct Effect-Size Measurement

As previously stated, the outcomes could be presented in two ways: dichotomous or continuous. Based on this, the effect size of the treatment

should be presented properly, using one or more of the following parameters [5, 17].

The risk ratio (RR): also defined as the relative risk, is the ratio of the risk of an event in the two groups. It is used for dichotomous outcomes and ranges from 0 to infinity. The RR describes the multiplication of the risk that occurs using the experimental intervention. For example, an RR of 5 for a treatment implies that events with treatment are five times more likely than events with the control treatment. Alternatively, we can say that the experimental treatment increases the risk of events by $100 \times (RR - 1)\% = 400\%$. Conversely, an RR of 0.20 is interpreted as the probability of an event with experimental treatment being a fifth of that with control treatment. Alternatively, this reduction could be expressed using the relative risk reduction (RRR) (see below). The interpretation of the clinical importance of a given RR should be made considering the typical risk of events with the control treatment, since an RR of 5 could correspond to a clinically important increase in events from 10 to 50%, or a small, less clinically important increase from 1 to 5%.

The relative risk reduction (RRR): is an alternative way of re-expressing the RR as a percentage of the reduction of the relative risk after the experimental treatment compared with the control treatment. For example, an RR of 0.20 is interpreted as the probability of an event with experimental treatment being a fifth of that with control treatment. Since the RRR is calculated as $100 \times (1 - RR)\%$, we can therefore say that the experimental treatment reduces the risk of events by 80%.

The odds ratio (OR): considering “odds” as the ratio between the probability that a particular event will occur and the probability that it will not occur, the OR is the ratio of the odds of an event in the two groups. It is used for dichotomous outcomes and ranges from 1 to infinity. The OR is more difficult to interpret, as it describes the multiplication of the odds of the outcome that occur using the experimental intervention. To understand what an OR means in terms of changes in the numbers of events, it is simplest to convert to it into an RR and then interpret the risk

ratio in the context of a typical control group risk. Attention should be paid to not misinterpreting RR and OR.

The risk difference (RD): is defined as the difference between the observed risks (proportions of individuals with the outcome of interest) between the two groups of experimental and control treatment. It is used for dichotomous outcomes and could be comprised between -1 and $+1$ (which means that it could be easily converted to per cent by multiplying it by 100). Like the RR, the clinical importance of an RD may depend on the underlying risk of events in the control treatment, since an RD of 0.05 (or 5%) may represent a small, clinically insignificant change from a risk of 55 to 60% or a proportionally much larger and potentially important change from 1 to 6%.

The number needed to treat (NNT): is defined as the expected number of people who need to receive the experimental treatment to obtain one additional person either incurring or avoiding the considered event. It is used for dichotomous outcomes and is always a positive number, usually rounded up to the nearest whole number. It is derived from the absolute value of $|RD|$ and is calculated as $1/|RD|$. An RD of 0.23 therefore corresponds to an NNT of 4.34, which is rounded up to 5, and it means that “it is expected that one additional (or less) patient will incur the event for every five participants receiving the experimental intervention rather than control”. Since the NNT gives an “expected value”, it does not imply that one additional event will necessarily occur in every group of “ n ” patients treated with the experimental procedure.

The mean difference (MD): is defined as the absolute difference between the mean value of the experimental and control group. It is used for continuous outcome measured with the same scale and estimates the amount by which the experimental intervention changes the outcome on average compared with the control.

The standardised mean difference (SMD): is defined as the size of the intervention effect in each study relative to the variability observed in the study. Since it is used when the same outcome is measured with different scales (e.g. knee function

measured with Lysholm, subjective IKDC or KOOS scores), the results should be statistically standardised to a uniform scale before being combined. Care should be taken when the scales have a different “direction” (e.g. a scale increase while the others decrease with disease severity); in this case, multiplying for -1 should be used when needed, to ensure that all the scales point in same direction.

47.5.3 Identify and Measure Heterogeneity

Heterogeneity is a term used to describe the variability of studies. Variability in the studied participants (e.g. males or females, old adult or adolescent patients), interventions (open or minimally invasive procedures, different grafts for reconstructions) and outcomes (objective or subjective) may be described as clinical heterogeneity, variability in study design and risk of bias may be described as methodological heterogeneity, while the variability in the treatment effects in the different studies is known as statistical heterogeneity [17, 21]. The last one is usually a consequence of clinical or methodological diversity, or both, between the studies. Studies with methodological flaws and small studies may overestimate treatment effects and can contribute to statistical heterogeneity. The statistical heterogeneity should therefore be examined and quantified using statistical tests, to implement measures to reduce the risk of bias [4]. The chi-square test (χ^2) assesses whether observed differences in results are compatible with chance alone: a low p -value provides evidence of heterogeneity in intervention effects. Since this measurement did not provide the “amount” of heterogeneity, quantification according to the Higgins I^2 statistic should be performed. This test produces a 0–100% value that represents the percentage of total variation across studies due to heterogeneity. According to the Cochrane Guidelines, this value is interpreted as follows: 0–40% heterogeneity might be not important, 30–60% may represent moderate heterogeneity, 50–90% may represent substantial heterogeneity and 75–100% considerable heterogeneity. However, there are

no empirically developed cut-off points to determine when there is too much heterogeneity to perform a meta-analysis, and it is left to the author’s discretion to determine whether a meta-analysis is appropriate, based on the results of the test of heterogeneity and clinical judgement.

47.5.4 Apply Strategies to Address High Heterogeneity: The Random Effect and Others

Since high heterogeneity implies dissimilarity in the studies, a meta-analysis should be conducted with caution, and several strategies should be implemented to consider this situation. These should be applied after checking whether data are correct and no errors have been made in data extraction [4, 5].

Perform random-effect meta-analysis: the two most frequently used models to conduct a meta-analysis are the fixed- (Mantel-Haenszel, inverse variance or Peto methods) [24] and random-effect (DerSimonian and Laird method) [6] models. The fixed-effect model investigates the question “What is the best estimate of the population effect size?”, assuming a common treatment effect and that the differences between studies are due to chance. This model should be used with low heterogeneity, as it gives greater weight to larger studies. On the other hand, the random-effect model investigates the question “What is the average treatment effect?”, assuming the distribution of the treatment effect along a range of values. This model should be used with high heterogeneity that cannot be explained, as it assumes that the treatment effect in the different studies is not identical due to clinical and methodological heterogeneity, and this model therefore gives less weight to larger studies. If the heterogeneity is not extreme, they frequently lead to similar results. On the other hand, they can produce different conclusions, and the use of fixed- or random-effect models should therefore be carefully considered based on the amount of heterogeneity, despite no guidelines existing in this direction (random effect is usually used with $I^2 > 50\%$) (Fig. 47.1).

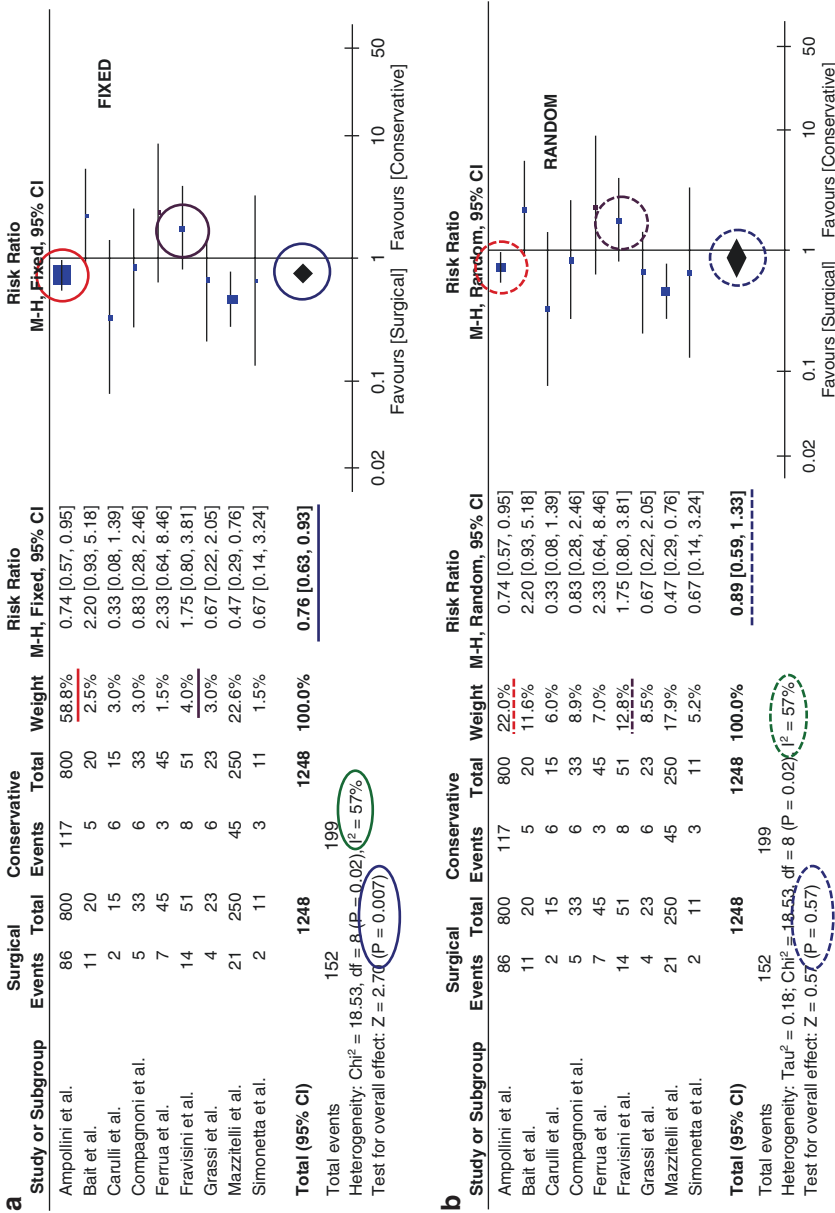


Fig. 47.1 Example of forest plots evaluating the risk ratio of treatment failure between the conservative and surgical treatment for a fictitious pathology. If a fixed-effect method was used, larger studies are represented with the largest squares (red circle) proportional to their weight (red line); similarly, the smallest studies have small squares (purple circle) and weight (purple line). In this specific case, the diamond of the overall effect (blue circle) crosses the central line and its confidence intervals do not contain the null value ($\text{RR} = 1$); it could therefore be assumed that there is a significant reduction in failure risk after surgical treatment ($P = 0.007$). However, the results should be interpreted with extreme caution due to the high and significant heterogeneity (green circle) (a). If a random-effect method is used to account for the high heterogeneity, the final result changes dramatically: the largest studies and smallest studies are given less weight (red dotted line and circle) and more weight (purple dotted line and circle), respectively, and this produces an enlargement of confidence intervals for the overall effect, which crosses the central line containing the null value ($\text{RR} = 1$). So, when random effect was used, we can affirm that there is no evidence of a significant effect of surgical treatment compared with conservative treatment in reducing failures ($P = 0.57$) (b)

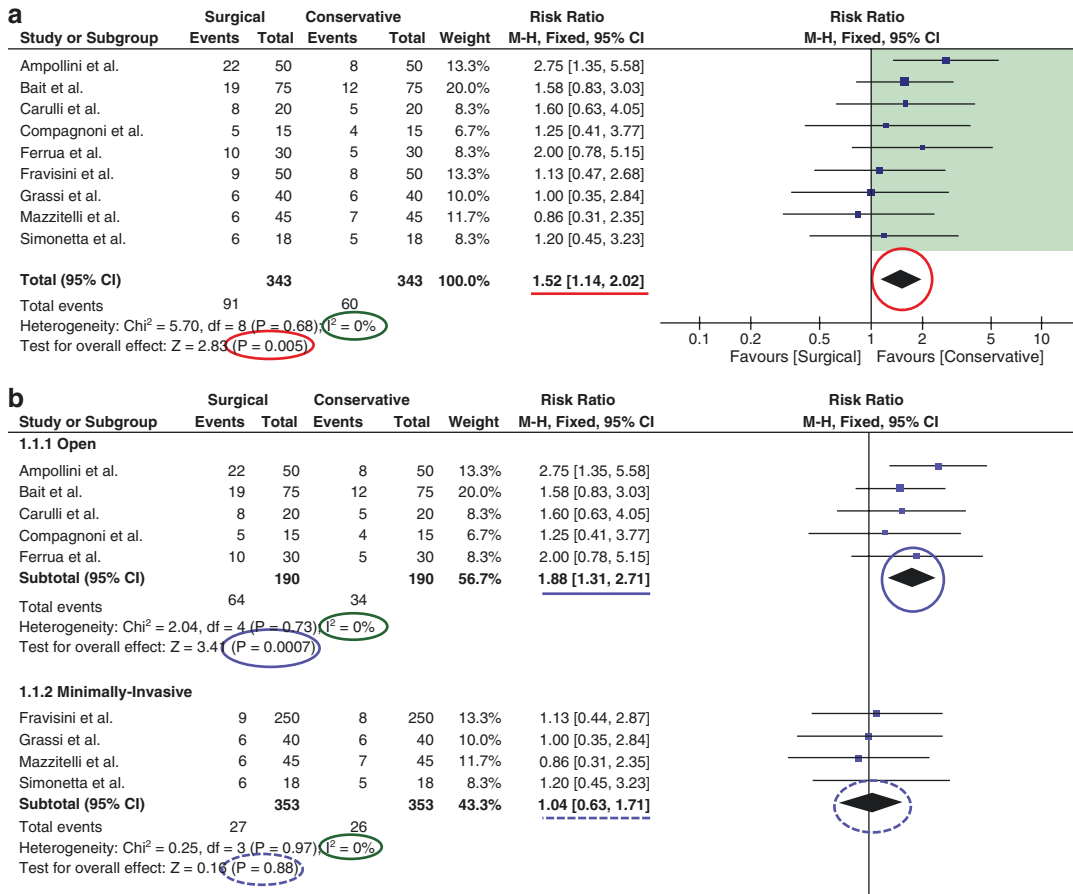


Fig. 47.2 Example of forest plot evaluating the risk ratio of complications after the conservative or surgical treatment for a fictitious pathology (a). In this case, it appears to be correct to use a fixed-effect model since the heterogeneity is null (green circle); this is confirmed by the concordance of the effect size of most of the studies, which lies on the right side of the forest plot. Since the confidence intervals do not contain the null value (RR = 1) (red circle and line), it could be suggested that there is evidence of an increased risk of complications after surgical treatment for the fictitious pathology. However, if a sub-

group analysis is performed separating open and minimally invasive surgery, the results differ from the main analysis (b). In the case of open surgery, the final overall effect is similar to the main analysis (blue line and circle); in the case of minimally invasive surgery, the confidence intervals of the overall effect (dotted blue line and circle) contain the null value (RR = 1), suggesting that there is no evidence of a difference in complications after minimally invasive surgery or conservative treatment for the fictitious pathology

Perform a subgroup analysis: subgroup analyses may be conducted as a means of investigating heterogeneous results or to answer specific questions about patient groups, types of intervention or types of study (Fig. 47.2). They can be performed for a subset of participants or a subset of studies, and they consider the meta-analysis results from each group separately. The non-overlap of the CI usually indicates statistical significance and a different effect of the

treatment within the subgroups, thereby explaining the heterogeneity to some extent. However, the subgroup analysis should be limited only to restricted cases, since it could increase the risk of type-II error due to the reduction of patient cohort size.

Perform a meta-regression: meta-regression is an extension of subgroup analyses that allows the investigation of the effect of continuous or categorical characteristics simultaneously. Its

role is like that of simple regression, where the outcome variable is the effect estimate (RR, OR, MD), and the explanatory variables, or covariates, are the study characteristics that might influence the size of the intervention effect. The regression coefficient obtained from a meta-regression will describe how the outcome variable changes with a unit increase in the covariate, while its statistical significance describes whether there is a linear relationship between them. It should be underlined that the characteristics to investigate (covariates) should be justified by biological and clinical hypotheses and should be the lowest number possible. One limitation of the meta-regression is that more than ten studies in the meta-analysis are generally required for its use.

Perform a sensitivity analysis: while the aim of the subgroup analysis is to estimate a treatment effect for a particular subgroup, the aim of the sensitivity analysis is to investigate whether the meta-analysis findings change based on different arbitrary or unclear decisions related to the meta-analysis process. The main decisions that can generate the need for a sensitivity analysis could be related to the eligibility criteria of the studies (e.g. study design or methodological issues), the data analysed (e.g. imputation of missing SD) or analysis methods (fixed or random effect, choice of effect-size measurement). In practical terms, the sensitivity analysis consists of the repetition of the meta-analysis, excluding the studies burdened by unclear or arbitrary decisions and in the informal comparison of the different ways the same thing is estimated. After the sensitivity analysis, when the overall conclusions are not affected by the different decisions made during the review process, more certainty can be assumed. On the other hand, if decisions are identified as influencing the findings, the results must be interpreted with an appropriate degree of caution if it is not possible to improve the process (Fig. 47.3). As different from the

subgroup analysis, the sensitivity analysis is not designed to estimate the effect of the intervention in the group excluded from the analysis, so its report should be produced with a summary table.

Change the measurement of effect: the choice of the measurements of effect size may affect the degree of heterogeneity; however, it is unclear whether the heterogeneity of intervention effect alone is a suitable criterion for choosing between the different measurements.

Exclude studies: since heterogeneity could be due to the presence of one or two outliers, studies with conflicting results compared with the rest could be excluded, as their exclusion could address the problem of heterogeneity. However, it is not appropriate to exclude a study based on its result since it may introduce a bias. It can therefore only be removed with confidence if there are obvious reasons. Unfortunately, there are no tests to determine the extent of clinical heterogeneity, and researchers must decide whether the studies contributing to a meta-analysis are similar enough clinically to make meta-analysis feasible. Refining inclusion criteria and excluding studies, even if this reduces heterogeneity, also decreases the total number of articles included on a topic. A sensitivity analysis is suggested to check whether the excluded study/studies could alter the meta-analysis results.

Do not perform a meta-analysis: if high heterogeneity cannot be addressed using the presented strategies, the investigator should consider whether the amount of heterogeneity is so large that the results of the meta-analysis are problematic. In this case, especially when there is inconsistency in the direction of the treatment effects that could make the use of an average value misleading, meta-analysis should be abandoned, and the evidence should be fairly expressed in a systematic review. Another and frequent reason to avoid meta-analytic pooling of data is when too few studies with no new findings are obtained after the systematic search.

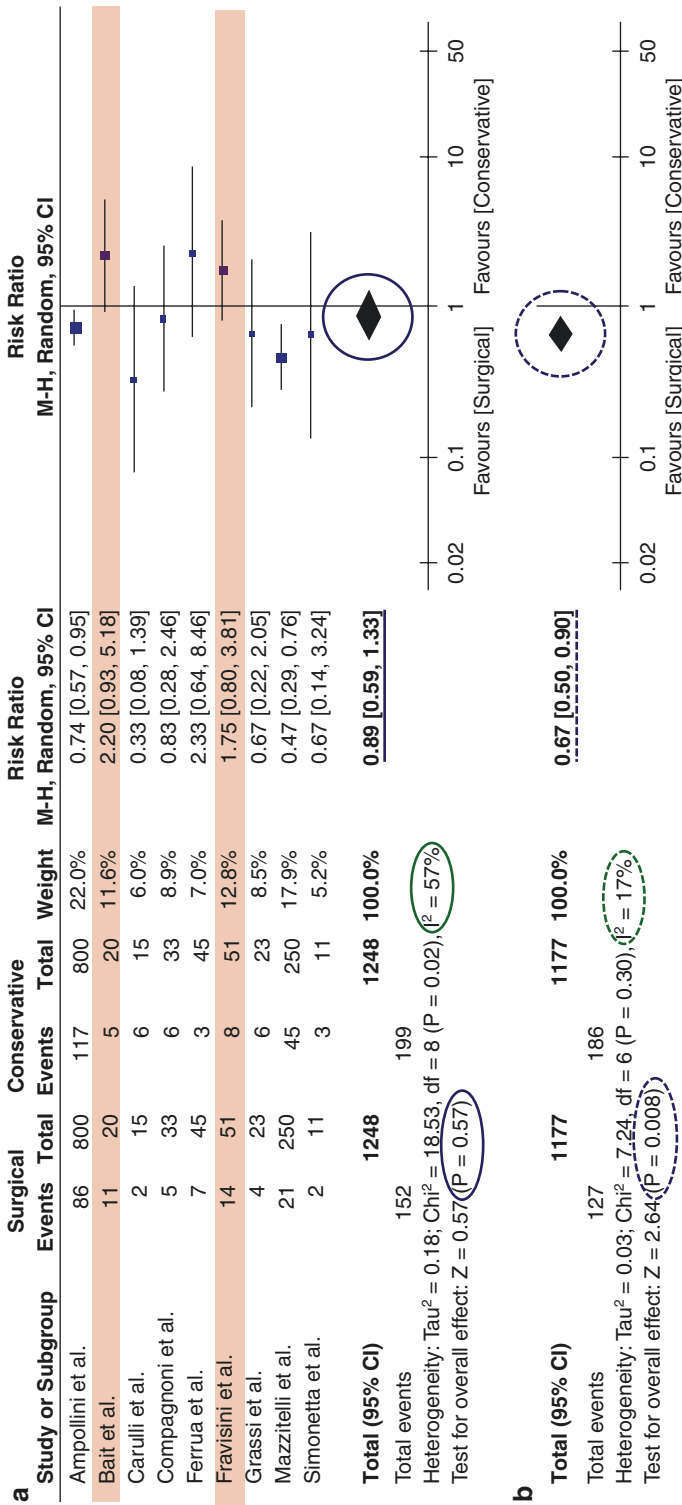


Fig. 47.3 Example of forest plots evaluating the risk ratio of treatment failure between conservative and surgical treatment for a fictitious pathology (a). After a random-effect meta-analysis due to the high heterogeneity (green circle), the final result is that there is no evidence of effect of surgical treatment in reducing the risk of failures, since the confidence intervals of the overall effect contain the null value (RR = 1). However, after risk-of-bias assessment, two studies with poor methodology and a high risk of bias have been found (Bait et al. and Fravisini et al.) and excluded through a sensitivity analysis (b). The results of the sensitivity analysis show a reduction in heterogeneity (green dotted circle) and the narrowing of the confidence intervals of the overall effect (blue dotted line and circle) that no longer contain the null value (RR = 1). In this case, the results of the meta-analysis should be interpreted with extreme caution since methodological issues and biases were able to affect the entirety of the overall effect of treatment

47.6 How to Present and Evaluate the Results

47.6.1 Prepare the Forest-Plot Graphic for the Main Outcomes

The result section of a meta-analysis should summarise the findings in a clear, logical order, explicitly addressing the objective of the review [11, 30]. The characteristics of methods, participants, intervention and outcomes should be reported in a narrative manner or with reference to tables [31]. On the other hand, the data analysis is better presented through the so-called “forest-plot” graphic; this is a simple, immediate and visually friendly method for describing the raw data, estimate and CI of the chosen effect measurement, the choice between fixed- or random-effect meta-analysis, the heterogeneity, the weight of each study and a test for the overall effect (Fig. 47.1). Forest plots should not be used when an outcome has only been investigated in a single study.

The measurement of the effect of each study included in the meta-analysis and in the forest plot is represented by a square, with the dimension proportional to its weight (based on sample size and the choice of a fixed- or random-effect model) and a horizontal line corresponding to its CI. Since the CI describes a range of values within which we can be reasonably sure that the true effect lies, a narrow CI indicates an effect size that is known precisely, while a very wide CI indicates that we

have little knowledge of the effect. The CI width for an individual study depends on the sample size, SD (for continuous outcomes) and risk of the event (dichotomous outcomes). When the CI crosses the central line (indicating an MD or SMD of 0 and an OR or RR of 1), it is possible that the experimental or control treatment has the same effect on the evaluated outcome. If the effect size of most of the included studies lies on the same side of the graphic, thus indicating a similar effect of the treatment, the overall heterogeneity is usually low.

The overall effect size of the meta-analysis is represented by a “diamond”. Its position indicates the value of the effect size, while its width indicates the CI. This width depends on the precision of the individual study estimates, the number of studies combined and the heterogeneity (in random-effect models, precision will decline with increasing heterogeneity). When the 95% CI for the effect of the meta-analysis does not cross the central line, it excludes the null value (MD or SMD of 0 and OR or RR of 1), and the *p*-value of the overall meta-analysis will therefore be <0.05. In this case, we can affirm that the observed effect is very unlikely to have arisen purely by chance and, as a result, there are differences in the effect of experimental and control interventions.

The forest plot, in a certain study design, could present the effect size of continuous or dichotomous outcomes from single-arm case series; in this case, we only have the estimation of pooled outcomes, without the comparison between two treatments (Fig. 47.4).

Studies	Estimate (95% C.I.)
Ampollini et al.	90 . 000 (87 . 189, 92 . 811)
Baitet al.	86 . 000 (84 . 131, 87 . 869)
Carulli et al.	88 . 000 (85 . 370, 90 . 630)
Compagnoni et al.	85 . 000 (77 . 301, 92 . 699)
Ferrua et al.	89 . 000 (86 . 093, 91 . 907)
Fravisini et al.	93 . 000 (90 . 977, 95 . 023)
Grassi et al.	91 . 000 (88 . 470, 93 . 530)
Mazzitelli et al.	90 . 000 (88 . 891, 91 . 109)
Simonetta et al.	95 . 000 (85 . 382, 104 . 618)
Overall (I ² =74.24 % ,P< 0.001)	89 . 547 (87 . 872, 91 . 222)

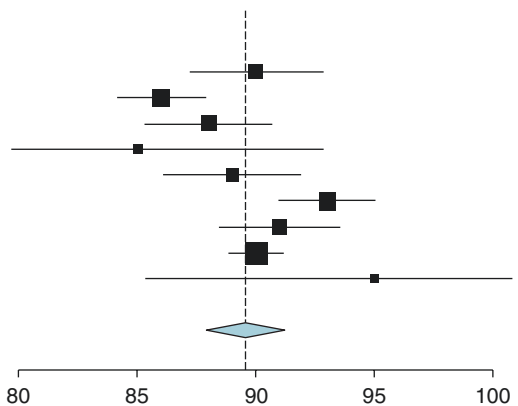


Fig. 47.4 Example of a forest plot of a continuous outcome from single-arm case series. In this case, no comparison between treatments is made, since a single

treatment is evaluated in the included studies; the main result is therefore the mean of the considered score or outcome, based on the weight of each study

47.6.2 Perform a Methodological Assessment and Bias Evaluation

The fundamental measurement to guarantee credibility in a meta-analysis is the evaluation of the methodology and bias of the included studies: this is a necessary step that should not be missed, because it could generate a misinterpretation of the results [12]. First, the level of evidence (Table 47.3) should be immediately and clearly reported. For single-arm case series, many methodological questionnaires are available, and one of the most frequently used is the Coleman Score [2] or its modifications [22] (Table 47.4). For non-randomised controlled studies, authors usually refer to the Newcastle-Ottawa Scale (NOS) [34] or its modifications (Table 47.5). For RCTs, there is also a vast choice of scores and checklists [27], with those obtained from the Consolidated Standards for Reporting Trials (CONSORT) guidelines regarded as some of the most authoritative [16] (Table 47.6). However, the indispensable action to ensure scientific strictness is the risk-of-bias evaluation, which is performed using the “Cochrane Risk of Bias Tool” [12].

A bias is defined as a systematic error (or a deviation from the truth) in results, and it can lead to an underestimation or overestimation of the true intervention effect. The types of bias considered in the Cochrane Risk of Bias Tool are:

- Selection bias: the systematic difference between the baseline characteristics of the groups

- Performance bias: the systematic difference between groups in the care that is provided
- Attrition bias: the systematic difference between groups in withdrawals from a study
- Detection bias: the systematic difference between groups in how outcomes are determined
- Reporting bias: the systematic difference between reported and unreported findings

According to these types of bias, seven domains are evaluated and rated as a low, unclear and high risk of bias.

- *Random sequence generation (selection bias)*: describe the method used to generate the allocation sequence in sufficient detail to allow an assessment of whether it should produce comparable groups.
- *Allocation concealment (selection bias)*: describe the method used to conceal the allocation sequence in sufficient detail to determine whether intervention allocations could have been foreseen in advance of, or during, enrolment.
- *Blinding of participants and personnel (performance bias)*: describe all the measures used, if any, to blind study participants and personnel to knowledge of the intervention a participant received. Provide any information relating to whether the intended blinding was effective. An evaluation should be made for each main outcome.
- *Blinding of outcome assessment (detection bias)*: describe all the measures used, if any, to blind outcome assessors to knowledge of the intervention a participant received. Provide any information relating to whether the intended blinding was effective. An evaluation should be made for each main outcome.
- *Incomplete outcome data (attrition bias)*: describe the completeness of outcome data for each main outcome, including attrition and exclusions from the analysis. State whether attrition and exclusions were reported, the numbers in each intervention group (compared with total randomised participants), reasons for attrition/exclusions where reported

Table 47.3 List of the five levels of evidence

Level of evidence for therapeutic studies	
Level I	Randomised controlled trial (RCT)
Level II	Prospective cohort studies (non-randomised comparative study)
Level III	Retrospective cohort study (non-randomised comparative study); case-control study
Level IV	Case series
Level V	Mechanism-based reasoning

Table 47.4 Modified Coleman Score used for orthopaedic case series (adapted from: Magnussen RA, Carey JL, Spindler KP. Does autograft choice determine intermediate-term outcome of ACL reconstruction? *Knee Surg Sports Traumatol Arthrosc.* 2011;19(3):462–472)

Modified Coleman Methodology Score		
Outcome	Option	Points
Part A: Only one score to be given for each section (total = 60)		
1. Study size: number of patients	>120	10
	81–120	7
	40–80	4
	<40 or not stated	0
2. Mean follow-up (years)	>6 years	5
	3–6 years	3
	<3 years, not stated, unclear	0
3. Percentage of patients with follow-up	>90%	5
	80–90%	3
	<80%	0
4. No. of interventions per group or separate outcomes should be reported	One procedure	10
	More than one procedure but consistent	5
	Among all patients in each group	
	Unclear or multiple interventions	0
5. Type of study	Randomised controlled trial	15
	Prospective cohort study	10
	Retrospective cohort study	5
6. Diagnostic certainty	In all	5
	In >80%	3
	In <80%, not stated, unclear	0
7. Description of surgical technique	Technique stated with details	5
	Technique named without elaboration	3
	Not stated, unclear	0
8. Description of postoperative rehabilitation	Well described	5
	Described without complete detail	3
	Protocol not reported	0
Part B: Scores could be given for each option in each of the three sections (total = 40)		
1. Outcome criteria	Outcome measurements clearly defined	2
	Timing of outcome assessment clear	2
	Use of outcome with good reliability	3
	Use of outcome with good sensitivity	3
2. Procedure for assessing outcomes	Subjects recruited	5
	Independent investigators	4
	Written assessment	3
	Patient-centred data collected	3
3. Description of subject selection process	Selection criteria reported and unbiased	5
	Recruitment rate reported >80%	5
	Eligible subjects not included in the	5
	Study satisfactorily accounted for	

The total score (max = 100) of the evaluated study is calculated as the sum of Parts A and B

Table 47.5 Modified Newcastle-Ottawa Scale for non-randomised studies (adapted from: <http://www.uphs.upenn.edu/cep/methods/Modified%20Newcastle-Ottawa.pdf>)

Question	Y\N
Study population:	
1. All study groups derived from similar source/reference populations?	
2. Attrition not significantly different across study groups?	
Study validity:	
3. The measurement of exposure is valid?	
4. The measurement of outcome is valid?	
5. Investigators blinded to end-point assessment?	
Confounders:	
6. Potential confounders identified (e.g. co-morbidities)	
7. Statistical adjustment for potential confounders made?	
8. Funding source(s) disclosed and no obvious conflict of interests?	
The scale does not require a calculation of a numeric score but can be used to map the study characteristics visually	

Table 47.6 Reporting quality scale for randomised controlled trials based on the Consolidated Standards for Reporting Trials (CONSORT) guidelines (adapted from: Huwiler-Müntener K, Juni P, Junker C, Egger M. Quality of reporting of randomized trials as a measure of methodologic quality. JAMA. 2002 Jun 5;287(21):2801–4)

Reporting quality scale based on 1996 CONSORT statement	
Question	Y\N
1. Does the title identify the study as a randomised controlled trial?	
2. Is the abstract presented in structured format?	
3. Are objectives stated?	
4. Is hypothesis stated?	
5. Is the study population described?	
6. Are the inclusion and exclusion criteria described?	
7. Are the interventions described?	
8. Are the outcome measurements described?	
9. Is a primary outcome specified?	
10. Is a minimum (clinically?) important difference for the primary outcome reported?	
11. Are power calculations described?	
12. Is the rationale for the statistical analyses explained?	
13. Are the methods for statistical analyses described?	
14. Are stopping rules described?	
15. Is the unit of randomisation described?	
16. Is the method used to generate the allocation schedule described?	
17. Is the method of allocation concealment described?	
18. Is the timing of assignment described?	
19. Is the method for separating those generating the allocation sequence from those assigning participants to groups described?	
20. Are the mechanisms of blinding described?	
21. Is the number of eligible patients reported?	
22. Is the number of randomised patients reported for each comparison group?	
23. Are prognostic variables by treatment and control group described?	
24. Is the number of patients receiving an intervention as allocated reported for each group?	
25. Is the number of patients analysed reported for each comparison group?	
26. Are withdrawals and drop-outs described for each comparison group?	
27. Are protocol deviations described for each comparison group?	
28. Is the estimated effect of intervention on primary and secondary outcomes stated, including measurements of precision?	
29. Are the results stated in absolute numbers?	
30. Are summary data and inferential statistics presented in sufficient detail to permit alternative analyses and replication?	
The total score (max = 30) of the evaluated study is calculated as the sum of the answer "Yes"	

Table 47.7 Strategies to identify high, low or unclear risk of bias for the main domains of the “Cochrane Risk of Bias Tool” (adapted from: Table 8.5c of Higgins JPT, Altman DG (editors). Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions. Chichester (UK): John Wiley & Sons, 2008)

Risk-of-bias assessment according to the “Cochrane Risk of Bias Tool”		
Selection	<i>Random sequence generation:</i> method used to generate the allocation sequence in sufficient detail to allow an assessment of whether it should produce comparable groups	
	<i>Low risk</i>	Randomisation with random number table, coin toss, computer generator, dice...
	<i>High risk</i>	Randomisation with date of birth, admission day, clinic record number, judgement of clinician...
	<i>Unclear risk</i>	Insufficient information on randomisation process
	<i>Allocation concealment:</i> method used to conceal the allocation sequence to avoid intervention allocations being foreseen in advance or during enrolment	
	<i>Low risk</i>	Participants could not foresee assignment due to central allocation, identical drug containers, opaque sealed envelopes
	<i>High risk</i>	Participants could foresee assignment due to open allocation schedule, alternation, unsealed envelopes
	<i>Unclear risk</i>	Insufficient information or concealment not described
Performance	<i>Blinding of participants and personnel:</i> measures used to blind study participants and personnel from knowing the intervention received	
	<i>Low risk</i>	Blinding or use of outcomes not influenced by lack of blinding
	<i>High risk</i>	No or incomplete blinding that could influence outcomes
	<i>Unclear risk</i>	Insufficient information to permit judgement
Detection	<i>Blinding of outcome assessment:</i> measures used to blind outcome assessors from knowledge of the intervention a participant received	
	<i>Low risk</i>	Blinding or use of outcomes not influenced by lack of blinding
	<i>High risk</i>	No or incomplete blinding that could influence outcomes
	<i>Unclear risk</i>	Insufficient information to permit judgement
Attrition	<i>Incomplete outcome data:</i> completeness of outcome data for each main outcome, including attrition and exclusion from the analysis	
	<i>Low risk</i>	No missing outcome data, or missing data balanced between groups and not related to outcomes
	<i>High risk</i>	Reason for missing data related to outcomes, or imbalance between intervention group
	<i>Unclear risk</i>	Insufficient reporting of exclusions to permit judgement
Reporting	<i>Selective reporting:</i> possibility of selective outcome reporting	
	<i>Low risk</i>	All the prespecified outcomes of interest have been reported
	<i>High risk</i>	Not all prespecified outcomes have been reported, or outcomes reported incompletely, or lack of key outcomes
	<i>Unclear risk</i>	Insufficient information to permit judgement
Other	<i>Other bias:</i> any important concerns about bias not addressed in the other domains in the tool	
	<i>Low risk</i>	The study appears to be free from other sources of bias
	<i>High risk</i>	Bias related to specific study design, or fraudulent, or other problems
	<i>Unclear risk</i>	Insufficient information to assess whether other biases exist

- and any re-inclusions in analyses performed by the review authors.
- *Selective reporting (reporting bias):* state how the possibility of selective outcome reporting was examined by the review authors and what was found.

- *Other (other bias):* state any important concerns about bias not addressed in the other domains in the tool.

According to the quality and methodology of the study, each domain should be rated (Table 47.7). The overall risk of bias should

Fact Box 47.4: Bias Included and Evaluated in the “Cochrane Risk of Bias Tool”

- **Selection bias:** the systematic difference between the baseline characteristics of the groups
- **Performance bias:** the systematic difference between groups in the care that is provided
- **Attrition bias:** the systematic difference between groups in withdrawals from a study
- **Detection bias:** the systematic difference between groups in how outcomes are determined
- **Reporting bias:** the systematic difference between reported and unreported findings

therefore be determined based on a low risk of bias for all key domains (low risk), a high risk of bias for one or more key domain (high risk) or an unclear risk for one or more key domain (unclear risk). Finally, the risk across studies could be defined as low if most information comes from studies with a low risk of bias, high if the proportion of information from studies with a high risk of bias is sufficient to affect the interpretation of the results and unclear if most information is from studies with a low or unclear risk of bias. To help the presentation of this information, a risk-of-bias summary (Fig. 47.5) and risk-of-bias graphs (Fig. 47.6) are extremely useful.

Other types of bias exist, due to imbalance in the dissemination of research findings due to the nature and direction of results. They are known as reporting biases [35] and can be:

- **Publication bias:** when the publication or non-publication of research findings depends on the nature and direction of the results; as an example, studies with negative results are often not published
- **Time-lag bias:** when the rapid or delayed publication of research findings depends on the nature and direction of the results

	Random sequence generation (selection bias)	Allocation concealment (selection bias)	Blinding of participants and personnel (performance bias)	Blinding of outcome assessment (detection bias)	Incomplete outcome data (attrition bias)	Selective reporting (reporting bias)	Other bias
Ampollini et al.	+	+		+	-	+	
Bait et al.	+	+	+	+	-		+
Carulli et al.	-	-	-	-	+	+	-
Compagnoni et al.			+	+	+	+	-
Ferrua et al.	+	-	+	+		+	
Fravisini et al.	+	+	-	-	+	+	
Grassi et al.	+	+			+	+	+
Mazzitelli et al.			-	-	+	+	
Simonetta et al.	+		+		+		-

Fig. 47.5 In this risk-of-bias summary, it is possible to have a visual presentation of the risk of each bias for all the included studies. The risk could be low (green plus), high (red minus), or unclear (no sign). In this specific table, it is possible to observe that the study by Carulli et al. is the most biased and the study by Grassi et al. is the one with the lowest risk of bias

- **Multiple publication bias (duplicate):** when the multiple or singular publication of research findings depends on the nature and direction of the results
- **Location bias:** when the publication of research findings in journals with different ease of access or levels of indexing in standard databases depends on the nature and direction of results
- **Citation bias:** when the citation or non-citation of research findings depends on the nature and direction of the results

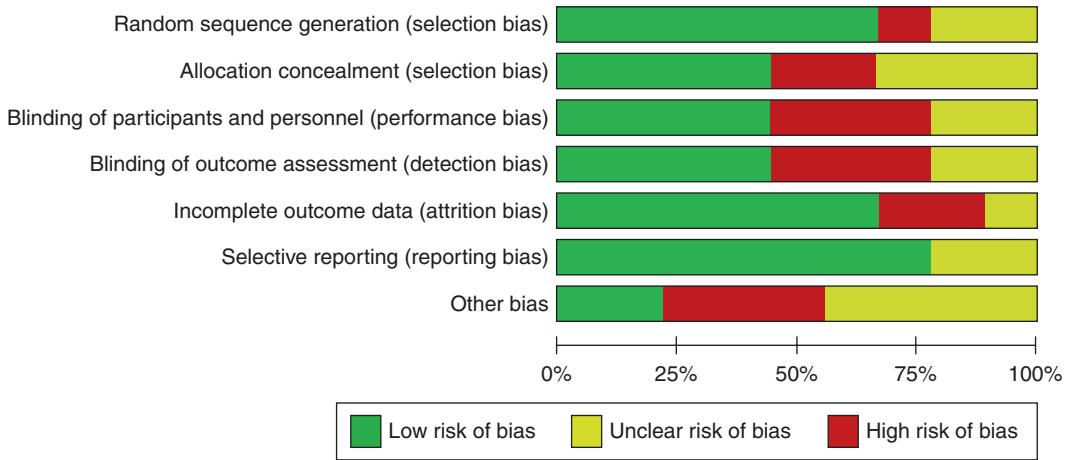


Fig. 47.6 In these risk-of-bias graphs, it is possible to see a visual presentation of the recurrent bias based on each domain. In this specific case, the “reporting bias” appears

to be the bias with a lower risk, while the “performance bias” and the “detection bias” appear to be those with a higher risk

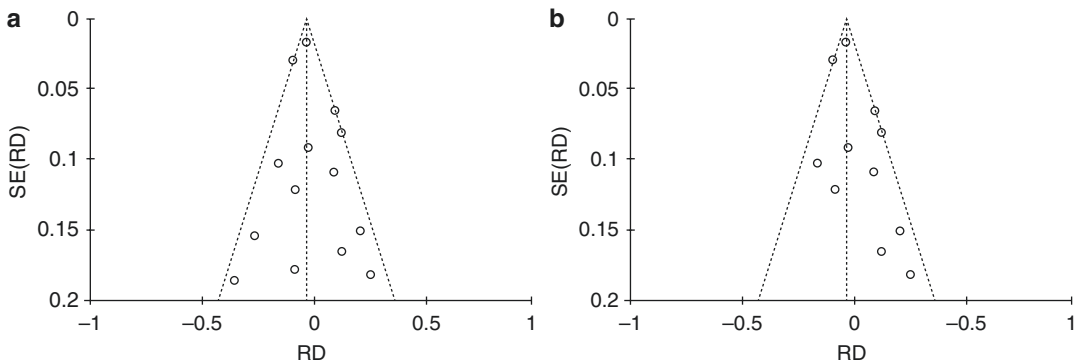


Fig. 47.7 In a funnel plot where the risk of bias is low, the symmetrical shape of an inverted funnel is seen (a). When the publication bias tends to exclude the publica-

tion of small studies without statistical significance, the funnel plot results are asymmetrical, with a gap in the bottom corner (bottom left corner, in this case) (b)

- **Language bias:** when the publication of research findings in a language other than English are sometimes be regarded as of secondary importance, while studies publishing positive results might also be more likely to publish in English

Outcome reporting bias: when the selective reporting of some outcomes but not others depends on the type of results found, if they are positive or negative or if they introduce new or repetitive findings.

One practical way to detect reporting bias is the use of the funnel plot graph [35]. This is a

simple scatter plot of the intervention effect estimates from individual studies against some measurement of each study’s size of precision; the effect estimates are plotted on the horizontal scale and the measurement of study size on the vertical axis (Fig. 47.7). As effect estimates from small studies scatter more widely at the bottom of the graph and those of larger studies are scattered more narrowly at the top of the graph, the plot should assume the shape of a symmetrical inverted funnel. If, due to publication bias, smaller studies without statistical significance remain unpublished, the plot will be asymmetrical with a gap in a bottom corner. In this case, the

meta-analysis will tend to overestimate the intervention effect. Apart from publication bias, asymmetry of the funnel plot could also be due to poor methodological quality, true heterogeneity, artefactual or chance. Funnel plot asymmetry should only be used if at least ten studies are included in the meta-analysis, when all the studies do not have similar sizes.

47.6.3 Correctly Approach and Evaluate Non-randomised Studies

Finally, a few words should be devoted to the meta-analysis of non-randomised controlled studies. Pooling together the results of non-randomised studies could be appropriate when they have a large effect; however, combining RCT and non-randomised studies is not recommended, as their results should be expected to differ systematically, resulting in increased heterogeneity [29].

Meta-analyses of non-randomised studies have greater potential bias, and their results should therefore be interpreted with caution. In fact, serious concerns could be related especially to the differences between people in different intervention groups (selection bias), caused by the lack of randomisation.

If both RCTs and non-randomised studies of an intervention are available and the author also wants to include a non-randomised study due to the small number of RCTs, they should be presented separately, or the findings of the non-randomised studies should be discussed in the final discussion with the meta-analysis findings.

47.7 How to Interpret Your Findings Critically

47.7.1 Summarise Your Main Findings

After the results have been correctly and clearly reported and methodology and bias adequately evaluated, the main findings of the meta-analysis can be critically interpreted.

For this purpose, “summary of findings” tables could be useful, since it presents the main findings in a simple format, providing key information on the quality of evidence, the magnitude of the effect of the interventions and the sum of available data on the main outcomes [31]. Six elements should be reported: a list of all important outcomes, a measurement of the typical burden of these outcomes, the absolute and relative magnitude of effect, numbers of participants and studies addressing these outcomes, a rating of the overall quality of evidence for each outcome and a space for comments. Special mention should be made of the quality of evidence, which is assessed through the Grades of Recommendation, Assessment, Development and Evaluation (GRADE) tool [1]. It describes the body of evidence as “High”, “Moderate”, “Low” or “Very Low”, based on the methodological quality, directness of evidence, heterogeneity, precision of effect estimates and risk of publication bias.

47.7.2 Pay Attention: What Is “Statistically Significant” Is Also “Clinically Significant”

When numerical results are going to be interpreted, attention should be paid to the 95% CI, because, if it is narrow, the effect size is known precisely, while, if it is wider, the uncertainty is greater. The CI and the *p*-value of the meta-analysis are strictly linked, as a value of <0.05 will exclude the null value (OR, RR of 1 or MD, SMD of 0) from the interval between the CIs, thus suggesting that the experimental treatment has an effect compared with the control treatment.

However, even if the findings are statistically significant, the clinical meaning of the benefit of the experimental treatment should be accurately weighted. When the treatment effect is measured with RR or RD, an interpretation of the clinical importance cannot be made without knowledge of the typical risk of events without treatment. In fact, a risk ratio of 0.75, for example, could correspond to a clinically important reduction in events from 80 to 60% or a small, less clinically important reduction from 4 to 3%. Conversely,

when dealing with continuous scales and mean differences, the proper minimum clinically important difference (MCID) of the considered outcome should be considered. Since the MCID represents the smallest change in a treatment outcome that a patient would identify as important, it is possible that a mean difference, despite being statistically significant, could be irrelevant from a clinical point of view (e.g. an MD of 4 points in the subjective IKDC, where the MCID is 11.5 points) [3, 8] (Table 47.8). Moreover, the minimum detectable change (MDC), which is the minimum amount of change in a patient’s score that ensures the change is not the result of measurement error, should be considered [3]. Recently, in a JBJS commentary, the superiority of double-bundle ACL reconstruction compared

with single-bundle demonstrated in a Level I RCT has been questioned, since a difference of less than 1 mm in an arthrometric evaluation and around two points in a subjective IKDC were not considered clinically meaningful, despite being statistically significant [23].

Furthermore, conclusions should not be drawn too quickly without performing an accurate evaluation of heterogeneity through subgroup or sensitivity analysis. For example, Soroceanu et al. [33] reported a relative risk of re-rupture of 0.4 in favour of surgical repair compared with conservative treatment in the case of Achilles tendon rupture. However, since the authors found a not negligible heterogeneity of 35%, they identified the item of “functional rehabilitation” as a cause of heterogeneity through meta-regression. So,

Table 47.8 The minimum detectable change (MDC) and minimum clinically important difference (MCID) of the main clinical scores used in knee surgery (adapted from: Collins NJ1, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS) and Tegner Activity Score (TAS). *Arthritis Care Res (Hoboken)*. 2011 Nov;63 Suppl 11:S208–28. doi: <https://doi.org/10.1002/acr.20632>)

The clinical scores most used for knee evaluation			
Score	Condition	MDC	MCID
Subjective IKDC	Injuries	8.8–15.6	6.3 (6 m) to 16.7 (12 m)
Subjective IKDC	Mixed pathologies	6.7	11.5 (sensitive) to 20.5 (specific)
KOOS pain	Injuries	6–6.1	–
KOOS symptoms	Injuries	5–8.5	–
KOOS ADL	Injuries	7–8	–
KOOS sport/rec	Injuries	5.8–12	–
KOOS Qol	Injuries	7–7.2	–
KOOS pain	OA	13.4	–
KOOS symptoms	OA	15.5	–
KOOS ADL	OA	15.4	–
KOOS sport/rec	OA	19.6	–
KOOS Qol	OA	21.1	–
Lysholm	Injuries	8.9–10.1	–
Lysholm	Mixed pathologies	–	–
Oxford Knee Scale	OA	6.1	–
WOMAC pain	OA	14.4–16.2	22.87 (TKR 6 m) to 27.98 (TKR 24 m)
WOMAC symptoms	OA	22.9–30.6	14.43 (TKR 6 m) to 21.35 (TKR 24 m)
WOMAC function	OA	10.6–15	19.01 (TKR 6 m) to 20.84 (TKR 24 m)
Tegner Activity Scale	Injuries	1	–
Tegner Activity Scale	OA	–	–

MDC minimum detectable change, MCID minimum clinically important difference

after performing a subgroup analysis separating patients undergoing functional or conventional rehabilitation, they found no difference in the re-rupture rate between surgical treatment and conservative treatment with functional rehabilitation. On the other hand, as Foster et al. [7] intended to include as many data as possible in their meta-analysis of irradiated vs. nonirradiated allografts for ACL reconstruction, they performed a sensitivity analysis to evaluate whether imputing SD would have influenced the final results. After performing an analysis of only the studies reporting SD, they repeated the analysis also adding those studies in which the SD was imputed as the mean value, reporting no substantial differences in the results. In their final evaluation, they therefore disclosed this issue and presented the data relative to all the studies, independently of the source of the SD.

47.7.3 Translate Your Findings into Clinics

The final difficulty in interpreting the meta-analysis results lies in applying the results to clinical practice [32]. It is important to correctly disclose whether the individual studies pooled in the meta-analysis can be generalised to a specific clinical scenario. This includes ensuring similar patient populations, interventions and outcomes of interest. For example, Jiang et al. [18] reported no differences in the rates of return to sport between patients undergoing surgical repair or non-surgical treatment after Achilles tendon rupture. However, since the RCTs included in the meta-analysis evaluated patients with a mean age of around 40 years, this recommendation should be applied with extreme caution in young, professional athletes.

Finally, attention should be paid when interpreting inconclusive or counter-intuitive results, which is one of the most common errors in scientific manuscripts. When there is inconclusive evidence, it is not appropriate to state that “there is evidence of no effect”. It is instead more appropriate to state that “there is no evidence of an effect”.

When results are instead counter-intuitive, clinical judgement based on experience, education and current practices will be needed to decipher the unexpected results. The decision to determine whether to accept the findings or question the statistical technique could be taken after looking back at the original articles, reassessing their inclusion and evaluating whether assumptions about the original research question are not lost when the studies are combined.

47.8 Conclusion: How to Prepare the Manuscript

The very last step in meta-analysis is to prepare a manuscript that is complete, essential, clear to the reader and suitable for publication in a peer-reviewed journal. First, the guideline of the target journal should be consulted to “tailor” the manuscript accordingly. Then, the PRISMA guidelines (Table 47.9) should be followed to fulfil the highest quality standard [25].

Title: should be concise and focused on the topic, identifying the paper as a meta-analysis.

Abstract: should be structured, including all the sections of the paper.

Introduction: should be short and focused on the topic, expressing the rationale of the meta-analysis, the purpose and the hypothesis.

Methods: should mention all the information regarding the databases used, the timing of the search, the keywords, the eligibility criteria, the methods of data extraction and the items evaluated. The statistical method used to combine the results, the bias evaluation and eventual sensitivity or subgroup analysis should be mentioned as well.

Results: should be clear and easy to understand. All included and excluded studies should be described in a flow diagram. It is recommended to present the data through forest plots and summary tables. The results of sensitivity or subgroup analysis and of bias evaluation should be provided in this section.

Discussion: should not be too long and should preferably focus on the main findings of the meta-analysis. The evidence should be sum-

Table 47.9 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) checklist as guideline for the final production of a meta-analysis (adapted from: Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. *BMJ*. 2009 Jul 21;339:b2700)

Preferred reporting items for systematic reviews and meta-analyses (PRISMA) checklist			
Section	Item	Description	Page
Title	1	Title	Identify the report as a systematic review, meta-analysis, or both
Abstract	2	Structured summary	Include, as applicable: background; objectives; data sources; eligibility criteria, participants and interventions; synthesis methods; results; limitations; conclusions
Introduction	3	Rationale	Describe the rationale for the review in the context of what is already known
	4	Objectives	Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes and study design (PICOS)
Methods	5	Protocol and registration	Indicate if a review protocol exists and, if available, provide registration information including registration number
	6	Eligibility criteria	Specify study characteristics and report characteristics used as criteria for eligibility, giving rationale
	7	Information sources	Describe all information sources and date last searched
	8	Search	Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated
	9	Study selection	State the process for selecting studies
	10	Data collection process	Describe the method of data extraction from reports
	11	Data items	List and define all variables for the data that were sought and any assumptions and simplifications made
	12	Risk of bias in individual studies	Describe the methods used for assessing risk of bias of individual studies and how this information is to be used in any data synthesis
	13	Summary measurements	State the principal summary measurements (e.g. risk ratio, difference in means)
	14	Synthesis of results	Describe the methods for handling data and combining results of studies
	15	Risk of bias across studies	Specify any assessment of risk of bias that may affect the cumulative evidence (e.g. publication bias)
	16	Additional analyses	Describe methods of additional analyses (e.g. sensitivity or subgroup analyses, meta-regression)
	Results	17	Study selection
18		Study characteristics	For each study, present characteristics of the data that were extracted
19		Risk of bias within studies	Present data on risk of bias of each study
20		Results of individual studies	For all outcomes considered, present for each study simple summary data for each intervention group and effect estimates with CI (forest plot)
21		Synthesis of results	Present results of each meta-analysis conducted, including confidence intervals and measurements of consistency.
22		Risk of bias across studies	Present results of any assessment of risk of bias across studies
23		Additional analysis	Give results of additional analyses, if performed (e.g. sensitivity or subgroup analyses, meta-regression)

(continued)

Table 47.9 (continued)

Preferred reporting items for systematic reviews and meta-analyses (PRISMA) checklist				
Section	Item	Description		Page
Discussion	24	Summary of evidence	Summarise the main findings including the strength of evidence for each main outcome	
	25	Limitations	Discuss limitations at study and outcome level and at review level	
	26	Conclusions	Provide a general interpretation of the results in the context of other evidence and implications for future research	
Funding	27	Funding	Describe sources of funding for the systematic review and other support; role of funders for the systematic review	

marised and discussed, together with the main limitations. The conclusions, which must be based exclusively on the findings without speculation, should delineate clinical and research implications.

Figures and tables: forest plots and funnel plots are very useful for result presentation, such as summary tables.

References: should be updated and formatted according to journal guidelines.

Funding: authors should always disclose any conflict of interests and eventual funding.

Clinical Vignette

After attending an International conference, you discover that the attention of Sports Medicine surgeons is on a new device to treat a specific type of ankle fractures, which has been introduced few years ago. You are aware of a couple of pilot RCT and, after performing a quick PubMed search, you find out at least three new RCTs published in the last year and a completed trial on clinicaltrials.gov, which is held by some of your overseas colleagues. Therefore, you plan to perform a systematic search to run a meta-analysis comparing this new device with the standard of care. With the help of your librarian and an orthopaedic resident in your hospital, you can define a broad and appropriate search strategy, analysing three databases and the website clinicaltrials.gov. Defining as inclusion criteria only RCT comparing the new device with the classic approach, you can find nine studies which

are pooled in a formal meta-analysis with the help of the biostatistician of your university. Due to the several differences in surgical procedure and patients' inclusion criteria, you opted to perform a more conservative statistical analysis using a random-effect model, considering also the high degree of statistical heterogeneity revealed with the I^2 test. Analysing the relative risk of complication and the mean differences of the main disease-specific scales, you find out a significant superiority of the new device. However, when you performed the bias evaluation, the lack of blinding of patients and clinicians raises some concerns due to the high risk of detection and performance bias. Overall, after the critical evaluation of results and bias, you agree with the chief of your clinic that the implementation of this new device in your clinical practice, with specific indications, could improve the quality of your treatments.

Take-Home Message

- Meta-analysis can be a powerful tool to combine results from studies with similar design and patient populations that are too small or underpowered individually.
- However, there are many potential threats that can limit the internal validity and real clinical impact of conclusions reported in a meta-analysis.
- An appropriate study question and design, the proper management of heterogeneity and the methodological evaluation of included studies with bias assessment are necessary to ensure the highest quality.

Appendix: Internet Links and Websites Useful for the Various Steps in Preparing a Meta-Analysis

Appendix: useful links	
Guides to meta-analyses	
Cochrane Handbook for Systematic Review of Interventions	http://handbook.cochrane.org/
PRISMA Guidelines for Systematic Reviews and Meta-Analyses	http://www.prisma-statement.org/
GRADE Handbook	http://gdt.guidelinedevelopment.org/app/handbook/handbook.html
Databases	
Cochrane Library	http://www.thecochranelibrary.com
PubMed	https://www.ncbi.nlm.nih.gov/pubmed/
Embase	http://store.elsevier.com/embase
Clinical Trials Database	https://clinicaltrials.gov/
Statistical software	
Cochrane RevMan	http://tech.cochrane.org/revman/download
OpenMetaAnalyst	http://www.cebm.brown.edu/openmeta/download.html
MedCalc	https://www.medcalc.org/download.php
Methodological evaluation	
Oxford Level of Evidence	http://www.cebm.net/oxford-centre-evidence-based-medicine-levels-evidence-march-2009/
JBJS Level of Evidence	http://jbjs.org/level-of-evidence
Newcastle–Ottawa Scale for Non-randomised Studies	http://www.ohri.ca/programs/clinical_epidemiology/oxford.asp
Modified Newcastle–Ottawa Scale	http://www.uphs.upenn.edu/cep/methods/Modified%20Newcastle-Ottawa.pdf
CONSORT checklist for randomised controlled trials (RCT)	http://www.consort-statement.org/
PEDRO scale for randomised controlled trials (RCT)	https://www.pedro.org.au/english/downloads/pedro-scale/
AMSTAR score for systematic reviews	https://amstar.ca/Amstar_Checklist.php
COSMIN guidelines for studies of measurement instrument	http://www.cosmin.nl/downloads.html (for studies of measurement instruments)

References

1. Brozek JL, Akl EA, Alonso-Coello P, Lang D, Jaeschke R, Williams JW, Phillips B, Lelgemann M, Lethaby A, Bousquet J, Guyatt GH, Schünemann HJ, GRADE Working Group. Grading quality of evidence and strength of recommendations in clinical practice guidelines. Part 1 of 3. An overview of the GRADE approach and grading quality of evidence about interventions. *Allergy*. 2009;64(5):669–77 . Review. <https://doi.org/10.1111/j.1398-9995.2009.01973.x>.
2. Coleman BD, Khan KM, Maffulli N, Cook JL, Wark JD. Studies of surgical outcome after patellar tendinopathy: clinical significance of methodological deficiencies and guidelines for future studies. *Victorian Institute of Sport Tendon Study Group. Scand J Med Sci Sports*. 2000;10:2–11.
3. Collins NJ, Misra D, Felson DT, Crossley KM, Roos EM. Measures of knee function: International Knee Documentation Committee (IKDC) Subjective Knee Evaluation Form, Knee Injury and Osteoarthritis Outcome Score (KOOS), Knee Injury and Osteoarthritis Outcome Score Physical Function Short Form (KOOS-PS), Knee Outcome Survey Activities of Daily Living Scale (KOS-ADL), Lysholm Knee Scoring Scale, Oxford Knee Score (OKS), Western Ontario and McMaster Universities Osteoarthritis Index (WOMAC), Activity Rating Scale (ARS), and Tegner Activity Score (TAS). *Arthritis Care Res (Hoboken)*. 2011;63(Suppl 11):S208–28. <https://doi.org/10.1002/acr.20632>.

4. Deeks JJ, Altman DG, Bradburn MJ. Statistical methods for examining heterogeneity and combining results from several studies in meta-analysis. In: Egger M, Davey Smith G, Altman DG, editors. *Systematic reviews in health care: meta-analysis in context*. 2nd ed. London: BMJ Publication Group; 2001.
5. Deeks JJ, Higgins JPT, Altman DG. Chapter 9: Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008.
6. DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials*. 1986;7:177–88.
7. Foster TE, Wolfe BL, Ryan S, Silvestri L, Kaye EK. Does the graft source really matter in the outcome of patients undergoing anterior cruciate ligament reconstruction? An evaluation of autograft versus allograft reconstruction results: a systematic review. *Am J Sports Med*. 2010;38(1):189–99. <https://doi.org/10.1177/0363546509356530>.
8. Grassi A, Ardern CL, Marcheggiani Muccioli GM, Neri MP, Marcacci M, Zaffagnini S. Does revision ACL reconstruction measure up to primary surgery? A meta-analysis comparing patient-reported and clinician-reported outcomes, and radiographic results. *Br J Sports Med*. 2016;50(12):716–24. <https://doi.org/10.1136/bjsports-2015-094948>.
9. Grassi A, Zaffagnini S, Marcheggiani Muccioli GM, Neri MP, Della Villa S, Marcacci M. After revision anterior cruciate ligament reconstruction, who returns to sport? A systematic review and meta-analysis. *Br J Sports Med*. 2015;49(20):1295–304. <https://doi.org/10.1136/bjsports-2014-094089>.
10. Grassi A, Zaffagnini S, Marcheggiani Muccioli GM, Roberti Di Sarsina T, Urrizola Barrientos F, Marcacci M. Revision anterior cruciate ligament reconstruction does not prevent progression in one out of five patients of osteoarthritis: a meta-analysis of prevalence and progression of osteoarthritis. *J ISAKOS*. 2016;1(1):16–24. <https://doi.org/10.1136/jisakos-2015-000029>.
11. Greco T, Zangrillo A, Biondi-Zoccai G, Landoni G. Meta-analysis: pitfalls and hints. *Heart Lung Vessel*. 2013;5(4):219–25. Review.
12. Higgins JPT, Altman DG. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008.
13. Higgins JPT, Deeks JJ. Chapter 7: Selecting studies and collecting data. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008.
14. Higgins JPT, Deeks JJ, Altman DG. Chapter 16: Special topics in statistics. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008.
15. Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008.
16. Huwiler-Müntener K, Jüni P, Junker C, Egger M. Quality of reporting of randomized trials as a measure of methodologic quality. *JAMA*. 2002;287(21):2801–4.
17. Israel H, Richter RR. A guide to understanding meta-analysis. *J Orthop Sports Phys Ther*. 2011;41(7):496–504. <https://doi.org/10.2519/jospt.2011.3333>.
18. Jiang N, Wang B, Chen A, Dong F, Yu B. Operative versus nonoperative treatment for acute Achilles tendon rupture: a meta-analysis based on current evidence. *Int Orthop*. 2012;36(4):765–73. <https://doi.org/10.1007/s00264-011-1431-3>.
19. Koretz RL, Lipman TO. Understanding systematic reviews and meta-analyses. *JPEN J Parenter Enteral Nutr*. 2016. pii: 0148607116661841.
20. Lefebvre C, Manheimer E, Glanville J. Chapter 6: Searching for studies. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008.
21. Lefavre KA, Slobogean GP. Understanding systematic reviews and meta-analyses in orthopaedics. *J Am Acad Orthop Surg*. 2013;21(4):245–55. Review. <https://doi.org/10.5435/JAAOS-21-04-245>.
22. Magnussen RA, Carey JL, Spindler KP. Does autograft choice determine intermediate-term outcome of ACL reconstruction? *Knee Surg Sports Traumatol Arthrosc*. 2011;19(3):462–72.
23. Marx RG. Anatomic double-bundle anterior cruciate ligament reconstruction was superior to conventional single-bundle reconstruction. *J Bone Joint Surg Am*. 2013;95(4):365. <https://doi.org/10.2106/JBJS.9504.ebo804>.
24. Mantel N, Haenszel W. Statistical aspects of the analysis of data from retrospective studies of disease. *J Natl Cancer Inst*. 1959;22:719–48.
25. Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;339:b2535. <https://doi.org/10.1136/bmj.b2535>. No abstract available.
26. O'Connor D, Green S, Higgins JPT. Chapter 5: Defining the review question and developing criteria for including studies. In: Higgins JPT, Green S, editors. *Cochrane handbook of systematic reviews of interventions*. Chichester: Wiley; 2008.
27. Olivo SA, Macedo LG, Gadotti IC, Fuentes J, Stanton T, Magee DJ. Scales to assess the quality of randomized controlled trials: a systematic review. *Phys Ther*. 2008;88(2):156–75.
28. Parlamas G, Hannon CP, Murawski CD, Smyth NA, Ma Y, Kerkhoffs GM, van Dijk CN, Karlsson J, Kennedy JG. Treatment of chronic syndesmotic injury: a systematic review and meta-analysis. *Knee Surg Sports Traumatol Arthrosc*. 2013;21(8):1931–9. Review. <https://doi.org/10.1007/s00167-013-2515-y>.
29. Reeves BC, Deeks JJ, Higgins JPT, Wells GA. Chapter 13: Including non-randomized studies. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008.

30. Russo MW. How to review a meta-analysis. *Gastroenterol Hepatol (N Y)*. 2007;3(8):637–42.
31. Schunemann HJ, Oxman AD, Higgins JPT, Vist GE, Glasziou P, Guyatt GH. Chapter 11: Presenting results and ‘summary of findings’ tables. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008.
32. Schunemann HJ, Oxman AD, Vist GE, Higgins JPT, Deeks JJ, Glasziou P, Guyatt GH. Chapter 12: Interpreting results and drawing conclusions. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008.
33. Soroceanu A, Sidhwa F, Aarabi S, Kaufman A, Glazebrook M. Surgical versus nonsurgical treatment of acute Achilles tendon rupture: a meta-analysis of randomized trials. *J Bone Joint Surg Am*. 2012;94(23):2136–43. <https://doi.org/10.2106/JBJS.K.00917>.
34. Stang A. Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *Eur J Epidemiol*. 2010;25(9):603–5. <https://doi.org/10.1007/s10654-010-9491-z>.
35. Sterne JAC, Egger M, Moher D. Chapter 10: Addressing reporting biases. In: Higgins JPT, Green S, editors. *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008.
36. Wright JG, Swiontkowski MF, Tolo VT. Meta-analyses and systematic reviews: new guidelines for JBJS. *J Bone Joint Surg Am*. 2012;94(17):1537. No abstract available.

A Practical Guide to Writing (and Understanding) a Scientific Paper: Clinical Studies

48

Riccardo Compagnoni, Alberto Grassi,
Stefano Zaffagnini, Corrado Bait,
Kristian Samuelsson, Alessandra Menon,
and Pietro Randelli

Fact Box 48.1

To become a medical writer, it is necessary to understand medical concepts and terminology, be familiar with relevant guidelines and the structure and content of specific documents, and, finally, have a good set of writing skills. A topic of interest, supported by current literature, should be identified and the work planned in an accurate way before starting the enrollment of the patients.

48.1 Introduction

Writing a scientific paper is a relevant part of the activities of medical doctors. Publishing and reviewing is becoming increasingly important for the progress of medical knowledge, offering the opportunity to share the results of the work that has been done or studying the work of other researchers. This is critical for the evolution of modern science, considering that the work of one scientist is based upon the results of others. Clinical outcomes can only be improved through research, education, and patient care. All these experiences are shared with the global community, primarily through peer-reviewed research papers and review articles [11, 15].

Two aspects are the most challenging. First the long time needed to obtain a paper of good quality and second the style of writing, generally regarded as less attractive when compared with surgical procedures. Some studies have attempted to analyze the most challenging topic at the beginning, identifying the cognitive burden, group support, and mentoring, the difficulty involved in distinguishing between content and structure, and the backward design of manuscripts as the most relevant [18]. When a medical doctor starts to write a paper, the motivation is crucial; however, the time that is needed to write without taking too much time from everyday clinical activities requires commitment. To become a medical writer, it is necessary to understand medical concepts and terminology, be

R. Compagnoni (✉) · A. Menon
1° Clinica Ortopedica, ASST Centro Specialistico
Ortopedico Traumatologico Gaetano Pini-CTO,
Milan, Italy

A. Grassi · S. Zaffagnini
Dipartimento Scienze Biomediche e Neuromotorie,
Università di Bologna, Bologna, Italy

Ila Clinica Ortopedica e Traumatologica, Istituto
Ortopedico Rizzoli, Bologna, Italy

C. Bait
Istituto Clinico Villa Aprica, Como, Italy

K. Samuelsson
Department of Orthopaedics, Institute of Clinical
Sciences, The Sahlgrenska Academy, University of
Gothenburg, Gothenburg, Sweden

P. Randelli
1° Clinica Ortopedica, ASST Centro Specialistico
Ortopedico Traumatologico Gaetano Pini-CTO,
Milan, Italy

Laboratorio di Biomeccanica ed Innovazione
Tecnologica, Istituto Ortopedico Rizzoli,
Bologna, Italy

familiar with relevant guidelines and the structure and content of specific documents, and, finally, have a good set of writing skills [19].

The aim of this paper is to provide tips acquired from authors attempting to help to write good-quality clinical papers without wasting their time and how to publish them in the appropriate journals. We have selected specific topics, which are analyzed in dedicated paragraphs. The different selections are how to choose the topic, find the current literature, analyze the data, structure the paper, write the paper, and handle references and some final tips on how to manage the submission process.

48.1.1 How to Choose the Topic of Your Work

Before beginning to enroll patients, to collect data, or to research the literature, the first step is to choose a topic of interest for your work [17, 23]. To do this, the best way is to ask your leader what is of interest, considering the trends in surgical procedures or the introduction of innovative techniques [21]. Once a topic has been identified, an accurate search of the literature, using the most common databases, can help the researcher to identify the actual “hot” topic in that specific field of research. It is necessary to know the timelines to follow, the data that should be collected, and the review/approval process to be followed.

48.1.2 Find Current Literature

Medical science is based on data available in the literature. The widespread availability of information due to the use of the Internet has given clinicians an opportunity to access the literature well. The most used database in which life sciences and biomedical papers are collected is MEDLINE (Medical Literature Analysis and Retrieval System Online, MEDLARS Online). PubMed is a free search engine primarily accessing the MEDLINE database of references and abstracts on life sciences and biomedical topics [8].

It is important to remember that PubMed should not be confused with PubMed Central, a

free digital archive of articles, accessible to anyone from anywhere via a basic web browser. The full text of all PubMed Central articles is free to read, with varying provisions for reuse.

To obtain more effective results, there are some tips to take into consideration in terms of using PubMed. The home page has a link to some tutorials, which will help the researcher to use the search engine correctly. MeSH (Medical Subject Headings) is the NLM (National Library of Medicine)-controlled vocabulary thesaurus used for indexing articles for PubMed. This vocabulary helps to identify the correct words that should be used in medical research and helps the authors to find the correct subject headings. It is very important to remember that, in MeSH, the subject headings are arranged in a hierarchy and a search for a descriptor will include all the descriptors in the hierarchy below the given one [8].

Regional health and medical databases have been compiled by the WHO (World Health Organization) to complement the internationally known bibliographic indices such as MEDLINE.

The regional medical indexes, published by or under the auspices of WHO Regional Offices, provide access to bibliographic information about the health material published locally. They thus add a further dimension to the retrieval of information from developed country-oriented databases [10].

The Cochrane Library is a collection of databases in medicine and other health-care specialties provided by Cochrane and other organizations. At its core is the collection of Cochrane Reviews, a database of systematic reviews and meta-analyses that summarize and interpret the results of medical research. The Cochrane Library aims to make the results of well-conducted controlled trials readily available and is a key resource in evidence-based medicine [6, 9].

48.1.3 How to Decide on the Kind of Clinical Paper

There are different types of scientific article, some of which require original research (primary literature) and some that are based on other published work (secondary literature). It is important to have

a clear idea about the different types of article that you can publish in a specific journal. Original research comprises studies based on clinical activities and is classified as primary literature. This group includes original treatment studies or observational studies. Treatment studies are mainly randomized, controlled, clinical trials (RCT) or adaptive controlled trials. Observational studies are cohort studies that are prospective, retrospective, case control, and cross-sectional and finally case reports. A review article surveys and summarizes previously published studies, rather than reporting new facts or analysis, and is for this reason called secondary literature.

One important step is to determine the level of evidence, i.e., a ranking system used to describe the strength of the results measured in a clinical trial or research study. The levels of evidence are an important component of evidence-based medicine (EBM), and understanding the levels and why they are assigned to publications and abstracts helps the reader to prioritize the available information. Many different ways of grading the level of evidence are described, and many journals assign a level to papers that they publish. Randomized, controlled trials are generally defined as level 1 of evidence, but some aspects have to be considered to assess the quality of the paper, including randomization, blinding, a description of the randomization and blinding process, and a description of the number of subjects who withdraw or drop out of the study; the confidence intervals relating to study estimates; and a description of the sample size calculation [5].

Although the goal is to improve the overall level of evidence in medical practice, this does not mean that all lower-level evidence should be discarded. Case series and case reports are important for hypothesis generation and can form the basis of more relevant studies [5, 13].

48.2 How to Write the Paper

Many techniques for writing a scientific paper in an effective way have been described [12, 16]. The formal Introduction, Methods, Results, and Discussion (IMRAD) structure of scientific papers was adopted in the 1980s. Nowadays,

Fact Box 48.2

Clinical studies are written following the IMRAD structure since the 1980s (Introduction, Methods, Results, and Discussion). The abstract section should emphasize new and important aspects of the study or observations, without overinterpreting findings. The introduction should describe the background of the study and should finish with a statement of the clear aim of the study. Methods should explain the structure of the study, the way in which the results were obtained, and all the steps relating to patient management, from enrollment to final evaluation.

IMRAD is the format encouraged for the text of observational and experimental studies by the “Uniform Requirements for Manuscripts Submitted to Biomedical Journals,” which has become the most important, widely accepted guide to writing, publishing, and editing in international biomedical publications [4]. The Uniform Requirements are released by the International Committee of Medical Journal Editors (ICMJE). Some types of articles, such as meta-analyses, may require different formats, while case reports, narrative reviews, and editorials may have less structured or even unstructured formats [15]. Detailed suggestions on how to write the specific parts of the paper are available online at the ICMJE website. The UK National Knowledge Service provided funding to start the EQUATOR (Enhancing the Quality and Transparency of Health Research) project. This initiative seeks to improve the reliability of medical publications by promoting the transparent, accurate reporting of health research [1]. In the present article, the techniques most used by the authors are reported. If writers have a specific journal in mind in which to publish their work, it is crucial to read the instructions to authors and remember that many journals have open access instructions for reviewers. This is very useful, because it lets the researchers know which reviewer is going to check their work. The language must be

correct and, if researchers do not write fluent English, it is suggested that they should use online services or mother tongue colleagues/copy editors. There are limits to word count, usually 3000–4000 words (different for different journals), and many papers are too long and may be rejected for that reason alone.

The following section will provide some tips for the correct writing of the different sections of a manuscript.

48.2.1 Abstract

The usual sections defined in a structured abstract are background/purpose, methods, results, and conclusions. The abstract should provide the background to the study and should state the aim, the basic procedures/methods (selection of study participants, settings, measurements, and analytical methods), the main findings (giving specific effect sizes and their statistical and clinical significance—but not repeat which statistical methods were used), and principal conclusions. The results section is the most important part of the abstract, and nothing should compromise its range and quality. It should emphasize new and important aspects of the study or observations, without overinterpreting findings. Most journals require abstracts that conform to a formal structure within a word count of 200–250 words. Even if the abstract is the first part of any article and the only section freely available in the most used search engine, it should be written after the paper is finished.

The ICMJE recommends that journals publish the clinical trial registration number at the end of the abstract [3, 7]. Level of Evidence is also frequently added at the end of the abstract.

48.2.2 Introduction

The scope of the introduction is to show what is already known on the subject of the paper and—more important—to introduce what is unknown to justify the structure of the study and what is intended to be examined.

Many reviewers regard the introduction as a section of little interest and general—and lengthy—considerations about the topic or common knowledge should be avoided.

The introduction should finish with a statement of the clear aim of the study, with primary and secondary outcomes considered, and the hypothesis of the investigation, at least for clinical studies.

48.2.3 Methods

The methods section is an exact description of the work done by researchers involved in the study and should explain the structure of the study, the way in which the results were obtained, and all the steps relating to patient management, from enrollment to final evaluation.

The first statements have to describe the research design, the clinical diagnosis of the patients recruited, and in most cases the setting of the study.

If the study compares the results of the treatment in different groups, these groups have to be described carefully. Randomization is of central importance in clinical trials because it reduces bias and represents a basis for ensuring the validity of data analysis using statistical testing. The generation of an unpredictable allocation sequence represents the first crucial element of randomization in a randomized, controlled trial [20]. The two fundamental characteristics of randomization are that researchers must be unable to predict the group to which a patient will be assigned until the patient is unambiguously registered in the study and that researchers are unable to change a patient's allocation once he/she has been randomized. Remember that randomization with a low-quality allocation sequence can result in a biased estimation of the treatment effect [14, 22].

The treatment given to patients has to be described in a detailed manner. The length of the study from enrollment to conclusion has to be reported. Outcome measurements using instruments or scoring systems should be described and explained. Care should be taken to select the appropriate outcomes for the specific disease or

intervention; the outcome should have good reliability, validity, and sensitivity. When patient-reported questionnaires are used, it is important to report whether patients were blinded to the treatment received in the event of a randomized trial, in order to minimize the bias. It is also important to report whether the investigators that are dedicated to objective evaluations (e.g., clinical findings, radiographic measurements) were blinded to patient treatment. The primary and secondary outcomes must be clearly specified.

A crucial and often neglected section in a scientific manuscript is the statistical section. The software used for the statistical computation should be reported. Before performing the statistical analysis, the data should be tested for normality (e.g., using the Kolmogorov-Smirnov test), since the normal or non-normal distribution influences both the way measurements are reported (mean \pm standard deviation for normal, median, and interquartile ranges for non-normal distribution) and the choice of statistical tests. When comparing two independent groups, the “independent sample *t*-test” and the “Mann-Whitney test” should be used for normally or non-normally distributed continuous measurements, respectively. When similar comparisons relate to the same group (e.g., pre- vs. posttreatment), the “dependent sample *t*-test” or the “Wilcoxon test” should be used analogously. If categorical variables (e.g., IKDC grade) are compared, the “chi-square test” should be used. If three or more groups are being compared, the “ANOVA” or “Kruskal-Wallis test” should be used in the event of normal or non-normal distribution, respectively. For correlation analysis, the Pearson or Spearman tests should be used when the distribution of the variables is normal or not normal.

In the case of RCTs, a sample size calculation is mandatory to establish the sample size, since there is a great risk that an underpowered study will suffer from a so-called beta error (type-2 error), thereby possibly missing the chance to show a statistically significant finding. The sample size calculation must be performed prior to the start of the study; there are commercially available programs that can easily be used for this purpose.

Fact Box 48.3

If there is a large volume of data, it is suggested that results should be reported in tables, while repetition in the text section must be avoided. The aim of the discussion section is to interpret and describe the significance of the study findings in the light of what was already known about the research problem being investigated and to explain the impact of the results on that specific topic.

48.2.4 Results

The results section is the most important and is often the only section that is of interest to a stressed reader. All data must be reported carefully, like the number of patients that completed the study, the dropout rates in the different groups, and eventually the dropout rate associated with the specific treatment. It must be remembered that most quality score questionnaires regard a dropout rate exceeding 20% as an indicator of a low-quality study. The primary outcome results have to be expressed with statistical data. Any negative findings or unexpected collateral effects have to be reported. Based on normal or non-normal distribution, the standard deviations and the interquartile ranges should always accompany the mean and median value, respectively. If the number of patients included in the study is limited, individual patient data can be reported in a table.

If there is a large volume of data, it is suggested that they should be reported in tables, while repetition in the text section must be avoided. However, a summary of the main and most relevant findings is useful to the reader in order to focus on the results. All details can be given in tables and repetitions must be avoided.

48.2.5 Discussion

The purpose of the discussion section is to explain what your results mean and what contribution your paper makes to the field of study [22].

The aim of the discussion section is to interpret and describe the significance of the study findings in the light of what was already known about the research problem being investigated and to explain the impact of the results on that specific topic. Each of the main results of the study should be analyzed and discussed, possibly in the same order as they are reported in the results section. The discussion section is based on interpretation, which is a subjective exercise. For this reason, the author must avoid overinterpreting the results of the study, one of the most frequent errors made. On the other hand, the discussion section is not a repetition of the introduction or methods sections. The discussion section has to acknowledge any limitations of the study, especially in terms of the methodology, and any alternative explanations of the findings [2]. A concluding take-home message can restate the answer one last time and/or indicate the importance of the work by stating implications, applications, or recommendations [24].

48.2.6 References

Authors should provide direct references to original research sources whenever possible. Secondary references should be avoided. Science is based on the results produced by the scientific community, and the method the community uses to share results is publication in scientific journals. This database of information has become larger in recent years, thanks to the use of the Internet. This is an important opportunity for researchers who can access a great deal of information about a specific topic. Every published paper cites previous articles supporting the statements and background of the research, and these articles are cited at the end of the paper. References should not be used by authors, editors, or peer reviewers to promote self-interest. One suggestion for researchers is to use only references useful for the article, supporting specific purposes. Considering that science is always evolving and the impact of journals is different, references must be updated and, if possible, they should attempt to cite articles in the most relevant journals [15]. It is not superfluous to mention that

references should be formatted and ordered according to the specific journal guidelines. All too often this is not adhered to.

48.3 How to Manage the Submission Process

The submission process is the last step in the preparation of the article. This step is often complex, due to the different submission manager platforms at different journals. One suggestion to help researchers to make it easier is to prepare all the materials before starting the submission, also collecting information on ethics committees and clinical trial registration. More and more journals nowadays require a disclosure of financial conflicts of interest, which should be collected from every author, possibly before submission, especially in the case of multicenter studies. Sometimes, minor mistakes such as line numbering or line spacing determine the need for revision and time loss.

One useful tip is to check the PDF generated by the submission system to check for errors or mistakes.

48.4 Conclusion

Becoming a good scientific writer requires a long learning curve. The authors of this short guide feel that, to obtain good results in writing, a passion for clinical practice and being interested in analyzing the results of the clinical work are crucial. Writing and sharing the results of clinical work with others means taking part in something greater, whose aim is to obtain better results in the treatment of the patients.

Good luck!

References

1. Altman DG, Simera I, Hoey J, Moher D, Schulz K. EQUATOR: reporting guidelines for health research. *Lancet*. 2008;371(9619):1149–50.
2. Annesley TM. The discussion section: your closing argument. *Clin Chem*. 2010;56(11):1671–4. <https://doi.org/10.1373/clinchem.2010.155358>.

3. Andrade C. How to write a good abstract for a scientific paper or conference presentation. *Indian J Psychiatry*. 2011;53(2):172–5. <https://doi.org/10.4103/0019-5545.82558>.
4. Barron JP. The uniform requirements for manuscripts submitted to biomedical journals recommended by the International Committee of Medical Journal Editors. *Chest*. 2006;129(4):1098–9.
5. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg*. 2011;128(1):305–10. <https://doi.org/10.1097/PRS.0b013e318219c171>.
6. <http://www.cochraneflibrary.com>.
7. <http://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html>.
8. <https://www.ncbi.nlm.nih.gov/pmc/>.
9. https://en.wikipedia.org/wiki/Cochrane_Library.
10. <http://www.who.int/library/country/regional/en/>.
11. Kallestinova ED. How to write your first research paper. *Yale J Biol Med*. 2011;84(3):181–90.
12. Liumbruno GM, Velati C, Pasqualetti P, Franchini M. How to write a scientific manuscript for publication. *Blood Transfus*. 2013;11(2):217–26.
13. Levin KA. Study design IV. Cohort studies. *Evid Based Dent*. 2006;7(2):51–2.
14. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet*. 1998;352(9128):609–13.
15. Nahata MC. Tips for writing and publishing an article. *Ann Pharmacother*. 2008;42(2):273–7. <https://doi.org/10.1345/aph.1K616>.
16. O'Connor TR, Holmquist GP. Algorithm for writing a scientific manuscript. *Biochem Mol Biol Educ*. 2009;37(6):344–8.
17. Randelli PS, et al. Needs and wishes from the arthroscopy community. In: Karahan M, Kerkhoffs G, Randelli P, Tuijthof G, editors. *Effective training of arthroscopic skills*. Berlin: Springer; 2015.
18. Shah J, Shah A, Pietrobon R. Scientific writing of novice researchers: what difficulties and encouragements do they encounter? *Acad Med*. 2009;84(4):511–6. <https://doi.org/10.1097/ACM.0b013e31819a8c3c>.
19. Sharma S. How to become a competent medical writer? *Perspect Clin Res*. 2010;1(1):33–7. PubMed PMID: 21829780.
20. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet*. 2002;359:515–9.
21. Tuijthof G, Cabitza F, Ragone V, Compagnoni R, Dutch Arthroscopy Society Teaching Committee, Randelli P. What arthroscopic skills need to be trained before continuing safe training in the operating room? *J Knee Surg*. 2017;30(7):718–24.
22. Vickers AJ. How to randomize. *J Soc Integr Oncol*. 2006;4(4):194–8.
23. Whitlock EP, Lopez SA, Chang S, et al. Identifying, selecting, and refining topics. In: *Methods guide for effectiveness and comparative effectiveness reviews* [Internet]. Rockville: Agency for Healthcare Research and Quality (US); 2008.
24. Zeiger M. *Essentials of writing biomedical research papers*. New York: McGraw-Hill; 2000. p. 176–219.



Reporting Complications in Orthopaedic Trials

49

S. Goldhahn, Norimasa Nakamura,
and J. Goldhahn

49.1 Background

Nobody likes to report complications in a clinical study. However, complications in orthopaedic trials are an essential source of information. They may terminate unsuccessful treatment strategies, help to identify potential for development and form the basis for shared decision-making with patients.

Nevertheless, complications implicate different things to different parties. For surgeons, they seem to cause trouble in the first instance. In addition, they impair any success rate, may need re-intervention, often require extensive communication with patients, sometimes lead to legal problems and are often associated with more problems and high costs. Given their perception as failure, it is not surprising that some surgeons tend to neglect them—especially in reporting. Other surgeons are more critical, document and report more complications. So far, it is up to the surgeon's understanding and awareness what one

regards as a complication. The great variability of reported complications for specific indications illustrates this fact. A survey among orthopaedic surgeons supports this observation by demonstrating different awareness levels of complications [5]. A standardized approach for the documentation, assessment and reporting of complications in orthopaedic trials is suggested.

49.2 Different Perspectives on Complications

For legal authorities, complications are the so-called adverse events that must be reported according to the guidelines of good clinical practice (GCP). They are interested in information, whether the complication, for instance, leads to death or another stay in hospital (serious adverse event) or whether they are device related [6]. Reported complications may lead to a stop of a study, implant withdrawal from the market or legal consequences.

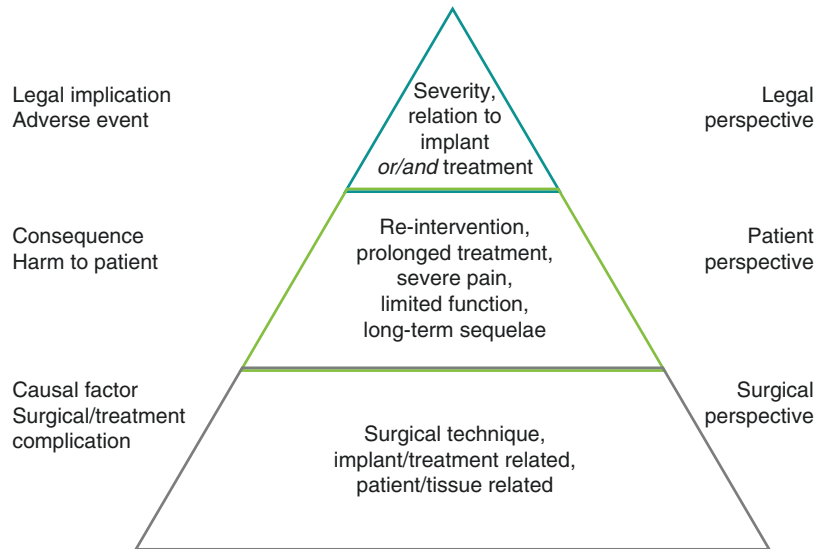
For patients, complications mean a decrease in quality of life in the first instance. A treatment may take longer than usual, may cause more pain than expected, may result in an inferior result and may lead to long-term sequelae. It could also result in a re-intervention to correct these conditions or to prevent long-term consequences. Primarily, patients are neither inter-

S. Goldhahn (✉)
Goldhahn GmbH, Baden, Switzerland
e-mail: goldhahn@goldhahn.swiss

N. Nakamura
Department of Rehabilitation Science, Osaka Health
Science University, Osaka, Japan
e-mail: norimasa.nakamura@ohsu.ac.jp

J. Goldhahn
Institute for Translational Medicine, ETH Zurich,
Zurich, Switzerland
e-mail: jgoldhahn@ethz.ch

Fig. 49.1 Hierarchy of complications. The pyramid illustrates the hierarchy from causal factor to patient harm until legal adverse event classification. This corresponds to the different perspectives at the right side of the diagram (Source: J. Karlsson et al. A practical guide to research: design, execution, and publication. *Arthroscopy*. 2011 April 27, 4 Suppl:S92)



ested in the surgeons' perspective nor in the legal perspective. They simply want to get function and quality of life re-established, and they regard everything as a complication that deviates from the normal course of healing and rehabilitation. In addition, they should get unbiased information about expected complication risk as a base for shared decision-making [1].

The outlined consequences demonstrate that it seems almost impossible to satisfy all perspectives at the same time. Therefore, a pragmatic approach is required that should acknowledge relevance. A severe complication may lead to a decrease in surgical reputation and/or a withdrawal of an implant with some financial consequences for the manufacturer. However, a patient may suffer from the consequences of a complication for the rest of his/her life or even die. Therefore, complications have the highest relevance for the person experiencing it. Consequently, definitions of complications should be patient-centred.

This leads to a hierarchical approach to complications (see Fig. 49.1). Whereas the surgical perspective is based on experience and always includes reasoning and causality, the patient perspective serves as a filter. Any event without any harm or consequences to the patients might not be considered as a complication.

On top of the hierarchy, the legal perspective determines the relation to any tested implant or treatment and classifies the severity according to established guidelines. It is mostly a subset of the complications that may matter to the patients as described above.

In accordance with the guidelines of the International Conference on Harmonization (ICH) [E2A and E6(R2)] Integrated Addendum to GCP guideline] and the ISO 14155:2011(E), a serious adverse event (SAE) is defined as any untoward medical occurrence that:

- Results in death
- Is life-threatening (note: the term life-threatening in the definition of "serious" refers to an event in which the patient was at risk of death at the time of the event; it does not refer to an event which hypothetically might have caused death if it were more severe)
- Requires inpatient hospitalization or prolongation of existing hospitalization
- Results in persistent or significant disability/incapacity
- Necessitates medical or surgical intervention to prevent permanent impairment to a body structure or a body function
- Leads to foetal distress, foetal death or congenital abnormality or birth defect

Case Vignette

In the treatment of an unstable trochanteric fracture using a dynamic hip screw, the screw was misplaced very close to the articular surface. The patient claims severe pain during weight bearing.

- *Surgical perspective:* The complication is defined as a screw cut-out. Possible causes can be initial misplacement (surgical technique) and/or poor bone quality (patient/tissue related).
- *Patient perspective:* The patient experiences severe pain and reduced function, may have long-term consequences if untreated or will face a re-intervention to prevent them.
- *Legal perspective:* The severity classification depends on the possible re-intervention. Possible relation to implant depends on the judgement of the surgeon, whether the malpositioning was related to poor surgical technique and/or device.

The case example demonstrates different issues: (1) The patient suffers under all circumstances regardless of the causing factor or the legal classification. (2) The surgeon can influence the classification of adverse events, e.g. by accepting poor functional outcome or neglecting re-intervention.

49.3 The Normal Course of Healing

If the patient perception of a complication is any deviation from the normal course of healing and rehabilitation, then a definition of “normal” is required. Healing of any tissue like bone, cartilage or tendon has a broad range depending on patient characteristics as well as on the specific intervention. For instance, time to fracture union

is not clearly defined and depends on many confounding variables and on the assessment method [2, 3, 8]. Therefore, thresholds are required that distinguish normal from pathological course of healing. The same is valid for pain and return to function.

Whereas a certain amount of pain caused by wound and tissue healing after a surgical intervention is associated with the normal course of healing, prolonged pain has another cause in most cases. The same is valid for return to function and activities of daily living. A certain improvement of function with a wide range is expected at given time points after intervention. However, complete loss of function or significantly lower function than expected and subsequently impaired activities of daily living should be considered as a complication.

Thus, for both pain and return to function, thresholds should be determined for the normal expected course of healing. Everything outside should be considered as a complication or the consequence of a complication. Pain and low function are often only the symptom of an underlying, often anatomical problem (e.g. articular step, valgus deformity). If patients report severe pain and/or limitation of function, it is necessary to search for the underlying problem.

49.4 Essentials of Complication Reporting

For each study, the normal course of healing and rehabilitation including an evidence-based range should be defined. This includes pain and functional status at each follow-up and healing of any investigated tissue such as cartilage or bone.

Anticipated complications/adverse events should be listed in all study protocols with clear and objective definitions along with appropriate scientific references.

- For each complication, a minimum set of information should be documented due to the regulations to allow clinically meaningful evaluation and reporting.
- In clinical research, these variables should be presented as a standard adverse event/complication case report form (CRF) that is adapted for each study.

It is of importance to quantify the standard complication rate known from the clinical literature, the common salvage procedures and the outcome that can be expected.

In Table 49.1, minimal requirements for record keeping of complications in clinical studies are listed. Investigators are asked to fill in one form for each complication; however more than one event may be recorded on the same form if they occurred simultaneously and were unambiguously causally related (e.g. an implant failure simultaneously with a loss of reduction).

Because complications occur as part of a complex chain of events, a clear distinction should be made between:

- The complication/adverse events themselves
- Their most likely causal trigger factors
- Their treatment (which could be no action)
- Their consequences or outcomes as illustrated in Fig. 49.2

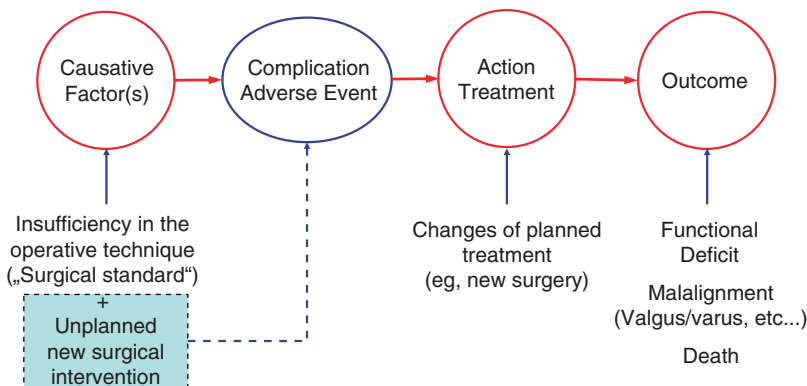


Fig. 49.2 Clear distinctions should be made between the complication and adverse event themselves, their most likely causal factors, their treatment (which could actually be no action) and their consequences or outcomes (Source:

Table 49.1 For each complication, a minimum set of information should be documented due to the regulations to allow clinically meaningful evaluation and reporting

Domain	Variables
Identification	1. Investigator’s name and phone number 2. Study name 3. Patient identification (trial number, initials, age, gender)
Treatment	4. The treatment number (if applicable, such as in a randomized clinical trial) 5. The name of the suspect medical product and date of treatment 6. Product serial number (in case of SADE)
Complication	7. Complication type 8. Date of occurrence or onset 9. Short description (open text field)
Action(s)	10. Subsequent action taken (e.g. operative)
Outcome(s)	11. Outcome of the complication at the time of reporting (or end of the study)
Assessment	12. Seriousness of the event 13. Most likely causative factor, e.g. relation to the surgical intervention or the implant used. We recommend using the four categories presented in this chapter

Note: This is the minimum information to be collected by means of an adverse event form/complication case report form (CRF) to be adapted for each study. Investigators are asked to fill in one form for each complication; however more than one event may be recorded on the same form if they occurred simultaneously and were unambiguously causally related
SADE serious adverse device effect

J. Karlsson et al. A practical guide to research: design, execution, and publication. Arthroscopy. 2011 April 27, 4 Suppl:S94

49.5 Classification of Complications

We propose the two main categories and subsequent two classes outlined in Table 49.2 for a classification of complications based on their most likely causative factor.

Of course, many cases remain where the causal relationship is the topic of debate. For instance, it is still not clear, whether an avascular head necrosis is the result of the surgical treatment of a humeral head fracture or would correspond to the normal course of disease.

However, careful planning combined with prospective definition of complications and their causal relationship increases the study quality. This planning phase may lead to an extensive list of anticipated complications [4] but helps to categorize complications prior to the study and will result in an unbiased complication analysis at the end of the study.

49.6 Follow-Up

- If an original complication record states that the complication was resolved or that the recovery process is completed (with or without damages), no further data are required.
- Alternatively, it is necessary to follow up the complication until it is resolved, in term of its treatment, outcome and assessment, and all new information must be documented.

In clinical studies, a follow-up adverse event/ complication CRF should be distributed to investigators to capture this information until complications are resolved or finally evaluated at the end of the studies.

49.7 Quality Control

Active monitoring and quality control are essential to avoid or limit under-reporting and misleading complication results. To favour completeness and correctness of documentation of complications, the following measures should be implemented in orthopaedic trials:

1. Source data verification during monitoring visits.
2. Active reporting: implement systematic assessment of any complication at each examination visit (e.g. using standard CRF or asking if another physician was visited other than for routine checks).
3. Incentive to report: facilitate simple recording process, and ensure anonymous reporting of complication statistics outside the involved clinics so that results cannot be traced back to the individual treating surgeon.
4. If necessary, additional information on putative events may be obtained from the patient’s family doctor.
5. Evaluation of reported complications by the study principal investigator, an independent experienced clinician or any specifically established Complication Review Board (CRB).

Table 49.2 Proposed classification of complications based on their most likely causative factor

Category	Class	Number	Example
Treatment related	Related to the surgical technique	1a	Malpositioning of screws, wrong procedure
	Related to the device/treatment	1b	Loosening of polyethylene glenoid due to wear
Patient related	Related to local tissue condition	2a	Cut-out of <i>correct placed</i> screw due to poor bone quality
	Related to overall patient condition, e.g. systemic	2b	Myocardial infarction

The final complication review should be conducted based on complication/AE forms, as well as additional diagnostics to complete the case. Complication data is reviewed for its clinical pertinence, classification, severity as well as relation to the investigated treatment or medical device. All changes and data corrections should be thoroughly justified and documented.

49.8 Analysis

A minimum set of complication analyses should be conducted in any study. However, it should be noted that if regulatory requirements oblige investigators to document all complications occurring during a study, only a specific clinically relevant subset may be analysed to answer a study objective. It is critical to clearly define which complications are included in such a subset and specify the period of observation (e.g. intraoperative, post-operative, follow-up periods) to allow appropriate interpretation of the results. In the context of prospective clinical investigations, the timing of observation for each patient starts with the initiation of treatment or primary surgery and ends at the end of the study. For the report of complications, complication risks can be calculated and presented as shown in Table 49.3.

Complication risks should be presented based on the number of patients experiencing complications and not on the total number of documented complications.

According to our experience, for many surgeons an event that is unrelated to the treatment may not be considered as a complication and must therefore not be documented. In addition, clinicians may sometimes feel they do not need to document events that have no or limited consequences for the patients to avoid documentation overload. Nevertheless, harmonized standards for the conduct of clinical trials define complication as “any untoward medical occurrence” not necessarily

Table 49.3 Example of presentation of complication risks

Type of complication	<i>n</i>	Risk (%)	95% CI
Post-operative local implant/ bone complications	18	10.2	(6.1–15.6)
<i>Implant</i>			
Blade migration	1	0.6	(0.01–3.1)
Implant breakage	3	1.7	(0.35–4.9)
Cut-out	2	1.1	(0.14–4.0)
Other implant complications	2	1.1	(0.14–4.0)
<i>Bone/fracture</i>			
Loss of reduction	1	0.6	(0.01–3.1)
Neck shortening	8	4.5	(2.0–8.7)
Other bone complications	6	3.4	(1.3–7.2)

Number of patients $N = 177$

n —number of patients with *at least one* complication (it means the patient can have more than one complication, but for the risk calculation, the *number of patients* experiencing complication(s) is used)

Risk—number of patients having a specific complication divided by the number of patients being enrolled in the study

95% CI—95% binomial exact confidence interval

related to potential causal factors or severity [7]; even mild anticipated complications in the framework of clinical research require official reporting to authorities if their rate of occurrence is higher than what can be reasonably be expected in any study. While all complications must be documented from a regulatory viewpoint, the primary analysis can be focused on the patient-relevant complications.

Conducting an independent review of complications is important for the credibility of safety data. Complication rates in the literature are most often elusive [5]. In addition, they are likely underestimated when documented by the inventor(s) of any surgical techniques. Despite all efforts at standardization, the assessment and reporting of complications will always require clinical judgement and therefore remain partly subjective. A Complication Review Board (CRB) can address such limitation, and can be established also for single-centre studies. A CRB can, for instance, consist of two to four orthopaedic surgeons (at least one of them should not be involved in the study), a radiologist and a methodologist. It is to be distinguished from any Data Monitoring Committee (DMC) [9] established as part of large multicentre studies; while the

CRB is set to control the relevance and integrity of the complication records, the DMC is set to review the occurrence of complications (i.e. assess the validated data and decide on the continuation of a study). The primary role of the CRB as proposed is to perform a quality control and consolidate complication data before their analyses.

At the end of a clinical study, complications/adverse events should be assessed and discussed by a Complication Review Board in a complication assessment meeting. The according complication case report forms, additional documentary material and all images of the patients should be available for such a meeting.

Take-Home Messages

- For each orthopaedic study, the normal course of healing including an evidence-based range should be defined.
- Anticipated complications/adverse events should be defined prior to study start together with a minimum set of documentation.
- It is necessary to follow up a complication until it is resolved.
- An independent Complication Review Board should review and analyse all information about potential orthopaedic complications.

References

1. Carlesso LC, MacDermid JC, Santaguida LP. Standardization of adverse event terminology and reporting in orthopaedic physical therapy: application to the cervical spine. *J Orthop Sports Phys Ther.* 2010;40(8):455–63.
2. Corrales LA, Morshed S, Bhandari M, Miclau T III. Variability in the assessment of fracture-healing in orthopaedic trauma studies. *J Bone Joint Surg Am.* 2008;90(9):1862–8.
3. Davis BJ, Roberts PJ, Moorcroft CI, Brown MF, Thomas PB, Wade RH. Reliability of radiographs in defining union of internally fixed fractures. *Injury.* 2004;35(6):557–61.
4. Goldhahn S, Kralinger F, Rikli D, Marent M, Goldhahn J. Does osteoporosis increase complication risk in surgical fracture treatment? A protocol combining new endpoints for two prospective multicentre open cohort studies. *BMC Musculoskelet Disord.* 2010;11:256.
5. Goldhahn S, Sawaguchi T, Audige L, Mundi R, Hanson B, Bhandari M, Goldhahn J. Complication reporting in orthopaedic trials. A systematic review of randomized controlled trials. *J Bone Joint Surg Am.* 2009;91(8):1847–53.
6. Hutchinson D, editor. *The Trial Investigator's GCP Handbook: a practical guide to ICH requirements.* Richmond: Brookwood Medical Publications; 1997.
7. ISO. ISO_14155:2011(E). *Clinical investigation of medical devices for human subjects - Good clinical practice.* International Standard Organization; 2011. p. 1–66.
8. Morshed S, Corrales L, Genant H, Miclau T III. Outcome assessment in clinical trials of fracture-healing. *J Bone Joint Surg Am.* 2008;90(Suppl 1):62–7.
9. USFDA. *Guidance for clinical trial sponsors. Establishment and operation of clinical trial data monitoring committees.* CBER; 2006. p. 1–34.



Understanding and Addressing Regulatory Concerns in Research

50

Jason L. Koh, Denise Gottfried, Daniel R. Lee,
and Sandra Navarrete

50.1 Introduction

Addressing regulatory concerns is important when any clinical trial is performed and especially if a new drug or device is being tested. Typically, these regulations are intended to protect the health and safety of human or animal subjects and patients but also can be related to ethical concerns. These help determine the appropriate conduct of a trial. Regulatory requirements typically depend on the known levels of safety and efficacy of a drug or device and also on the subjects being tested. Extensive requirements need to be met for testing on human or animal subjects. Significant training and an approved protocol are necessary before performing research on humans or animals.

With device testing, the regulatory burden depends on the classification of the device. An entirely novel device will typically have to go through a government-regulated multistage pre-market approval process. The development of the

appropriate protocols can be quite expensive and laborious involving repeated meetings with regulatory officials. The contact of the trial must be performed with respect to strict compliance with the protocol and can be subject to audit.

Other devices that are largely similar to previously approved or predicate devices can be assigned a different classification, and the regulatory requirements may be much less.

New drugs or materials also face different regulatory burdens depending on their classification. An entirely new drug typically has to go through a multiphase approval process, including demonstration of safety and efficacy. Even an existing drug that is seeking additional indications for use will have to go through a formal approval process.

Extensive regulation also guides the use of cell therapies and tissue. In the United States and in many other countries, the use of more than minimally manipulated cell therapies requires approval through an extensive formal regulatory process. This has contributed to the relatively lower amount of trials in such areas as stem cell research in the United States when compared to some other nations.

Ethical reasons may also drive regulation. Embryonic stem cell research in the United States has been significantly curtailed by policy that prohibited federal funding of such research due to ethical reasons related to the embryo source of the cells. Many countries ban aspects of human cloning, typically citing “human dignity” as the reason.

J. L. Koh (✉)
Department of Orthopaedic Surgery, NorthShore
University HealthSystem, Evanston, IL, USA

University of Chicago, Chicago, IL, USA

D. Gottfried
TekTeam, LLC, Palo Alto, CA, USA

D. R. Lee
Scaffold Biologics, Inc., San Antonio, TX, USA

S. Navarrete
Independent Consultant, Austin, TX, USA

The unregulated use of medical technologies can result in significant harm to patients as well as poor research. Failure by clinicians or researchers to comply with regulatory requirements can result in significant penalties, including being banned from performing certain research or clinical activities, civil liability, or criminal punishment including fines or even imprisonment. Therefore it is critical to be aware of the different regulatory issues regarding research and to be meticulously compliant.

The burden of meeting regulations may be quite high and can limit the speed of progress in research. There is a balance between the goals of protecting the public from inadequately tested drugs or devices and permitting the expeditious approval of new therapies that are of benefit to patients. The kind and type of regulation can vary related to politics, societal needs, or different weightings of ethical considerations.

50.2 Good Clinical Practice

The new researchers must be aware of their moral and ethical obligation to conduct research in a manner that assures the public that the rights, safety, and well-being of human research participants (i.e., subjects) are protected.

Good Clinical Practice (GCP) is defined by the International Conference on Harmonisation (ICH) as an international ethical and scientific quality standard for designing, conducting, recording, and reporting trials that involve the participation of human subjects. Compliance with this standard provides public assurance that the rights, safety, and well-being of trial subjects

are protected, consistent with the principles that have their origin in the Declaration of Helsinki, and that the clinical trial data are credible. The objective of the ICH GCP Guideline is to provide a unified standard for the European Union (EU), Japan, and the United States to facilitate the mutual acceptance of clinical data by the regulatory authorities in these jurisdictions. The guideline was developed with consideration of the current good clinical practices of the European Union, Japan, and the United States, as well as those of Australia, Canada, the Nordic countries, and the World Health Organization (WHO) [1].

The ICH E6 (R1) [2, 3] guidelines were developed in 1996 to harmonize the requirements for registering medicines in Europe, Japan, and the United States. Since they are internationally recognized, they permit all clinical trial evidence from one country to be accepted by another country. The general principles of ICH expand on the Declaration of Helsinki which are a set of ethical principles applicable to human research developed by the World Medical Association (WMA).

The ICH GCP E6 (R2) [4, 5] was finalized in 2016 and addresses the increased complexity of clinical trials and electronic data recording and reporting. The new draft is the biggest revision of the guideline in over 20 years.

A new clinical researcher should be aware of the terms sponsor and investigator to understand his/her role and responsibilities. Clinical trials may be initiated by a sponsor, investigator, or someone serving as both. Clinical trials involve several teams responsible for specific roles. The investigator may delegate certain roles to his/her team, but is responsible for the conduct of the clinical trial.

Fact Box 50.1: ICH Guidance Is Divided into Four Categories, and ICH Topic Codes Are Assigned According to These Categories [1]

Q: quality	S: safety	E: efficacy	M: multidisciplinary
Chemical and pharmaceutical quality of a drug (stability, validation, impurity testing)	Safety of the medicinal product (toxicology, carcinogenicity, genotoxicity)	Clinical studies in human subjects (dose response, GCP, trial design, conduct, analysis, AE reporting)	Issues that do not fall in the other categories (MedDRA—standardized medical coding for adverse events)

Fact Box 50.2**The 13 principles of ICH GCP [2]**

1. Clinical trials should be conducted in accordance with the ethical principles that have their origin in the Declaration of Helsinki, and that are consistent with GCP and the applicable regulatory requirement(s)
2. Before a trial is initiated, foreseeable risks and inconveniences should be weighed against the anticipated benefit for the individual trial subject and society. A trial should be initiated and continued only if the anticipated benefits justify the risks
3. The rights, safety, and well-being of the trial subjects are the most important considerations and should prevail over interests of science and society
4. The available nonclinical and clinical information on an investigational product should be adequate to support the proposed clinical trial
5. Clinical trials should be scientifically sound, and described in a clear, detailed protocol
6. A trial should be conducted in compliance with the protocol that has received prior institutional review board (IRB)/independent ethics committee (IEC) approval/favourable opinion
7. The medical care given to, and medical decisions made on behalf of, subjects should always be the responsibility of a qualified physician or, when appropriate, of a qualified dentist
8. Each individual involved in conducting a trial should be qualified by education, training, and experience to perform his or her respective task(s)
9. Freely given informed consent should be obtained from every subject prior to clinical trial participation
10. All clinical trial information should be recorded, handled, and stored in a way that allows its accurate reporting, interpretation and verification
11. The confidentiality of records that could identify subjects should be protected, respecting the privacy and confidentiality rules in accordance with the applicable regulatory requirement(s)
12. Investigational products should be manufactured, handled, and stored in accordance with applicable good manufacturing practice (GMP). They should be used in accordance with the approved protocol
13. Systems with procedures that assure the quality of every aspect of the trial should be implemented

Fact Box 50.3: The ICH GCP E6 (R2) Addendum Introduced 26 New Items Covering 3 Main Areas of Clinical Research: Data Management and Sponsor and Investigator Responsibilities [6, 7]

Number of amended items	Name of the amended section	Number and name of the amendment
1	Introduction	NA
4	Glossary	<ul style="list-style-type: none"> • Certified Copy 1.11.1 • Monitoring Plan 1.38.1 • Monitoring Report 1.39 • Validation of computerized systems 1.60.1
1	The principles of ICH GCP	<ul style="list-style-type: none"> • The principles of ICH GCP 2.10
3	Investigator	<ul style="list-style-type: none"> • Adequate Resources 4.2.5, 4.2.6 • Records and Reports 4.9.0
16	Sponsor	<ul style="list-style-type: none"> • Quality Management 5.0, 5.0.1, 5.0.2, 5.0.3, 5.0.4, 5.0.5, 5.0.6, 5.0.7 • Contract Research Organization (CRO) 5.2.1, 5.2.2 • Trial Management, Data Handling, and Record Keeping 5.5.3(b), 5.5.3 (h) • Monitoring 5.18.3, 5.18.6(e), 5.18.7 • Noncompliance 5.20.1
1	Essential documents for the conduct of a clinical trial	<ul style="list-style-type: none"> • Essential documents for the conduct of Clinical Trial 8.1

ICH GCP defines an investigator as a person responsible for the conduct of the clinical trial at a trial site. If a trial is conducted by a team of individuals at a trial site, the investigator is the responsible leader of the team and may be called the principal investigator [4, 5].

ICH GCP defines sub-investigator as any individual member of the clinical trial team designated and supervised by the investigator at a trial site to perform critical trial-related procedures and/or to make important trial-related decisions (e.g., associates, residents, research fellows) [4, 5].

ICH GCP defines sponsor as an individual, company, institution, or organization which takes responsibility for the initiation, management, and/or financing of a clinical trial [4, 5].

ICH GCP defines sponsor-investigator as an individual who both initiates and conducts, alone or with others, a clinical trial and under whose immediate direction the investigational product is administered to, dispensed to, or used by a subject. The term does not include any person other than an individual (e.g., it does not include a corporation or an agency). The obligations of a sponsor-investigator include both those of a sponsor and those of an investigator [4, 5].

It is important to note that corporate sponsors may ask investigators to complete GCP training prior to enrolling clinical research subjects as a means to ensure he/she understands their obligations during the study. Typically these training courses can be taken online, but they do have a cost associated with them.

It is worth mentioning that the ICH GCP Guideline clearly outlines essential documents for the conduct of a clinical trial. Essential documents are those documents which individually and collectively permit evaluation of the conduct of a trial and the quality of the data produced. These documents serve to demonstrate the compliance of the investigator and sponsor and monitor with the standards of Good Clinical Practice and with all applicable regulatory requirements [4, 5].

The Trial Master File (commonly referred to as the Regulatory File at the investigational site) houses the essential documents of the clinical trial and must be maintained with both the sponsor and investigator. Examples of essential

Fact Box 50.4: Examples of GCP Training Resources

- CITI Program online training course [8, 9].
- Barnett International Good Clinical Practice: A Question & Answer Reference Guide 2018 [10, 11]
- Clinical Device Group—Workshops on CD [12, 13].

documents that must be maintained by the investigator include the protocol and amendments, financial aspects of the trial, institutional review board (IRB) or independent ethics committee (IEC) approvals, decoding procedures for blinded trials, curriculum vitae for investigators and sub-investigators, and signature sheets to name a few. Since these files are subject to audit or inspection at any time, they must not be destroyed, and the duration they must be kept should be outlined at the start of each clinical investigation.

Compliance to GCP will help ensure scientific quality data when conducting trials that involve human participation because they provide the framework for ethical conduct and assure study documentation is complete. Regardless if the trial is a single-center or multicenter, small patient population or large, or whether it is sponsor initiated or investigator initiated, GCP should be followed by all parties involved in the clinical trial. Individual countries or institutions may have additional or similar GCP framework, so the investigator should refer to the appropriate legislation in which the research is being conducted. Following ICH GCPs facilitate regulatory approvals across countries, and in turn new treatment options become available to patients.

50.3 ICMJE: Clinical Trial Registration and Data Sharing

The International Committee of Medical Journal Editors (ICMJE) requires clinical research studies that began enrolling subjects on or before July 1,

Fact Box 50.5: Clinical Trial Registry Resources

www.clinicaltrials.gov [16]: [ClinicalTrials.gov](http://www.clinicaltrials.gov) is a resource provided by the US National Library of Medicine.

<http://www.isrctn.com> [17]: The ISRCTN registry is a primary clinical trial registry recognized by WHO and ICMJE that accepts all clinical research studies (whether proposed, ongoing, or completed), providing content validation and curation and the unique identification number necessary for publication.

<http://www.who.int/ictrp/network/primary/en/> [18]: Primary Registries in the WHO Registry Network meet specific criteria for content, quality and validity, accessibility, unique identification, technical capacity, and administration. Primary Registries meet the requirements of the ICMJE.

2005, will be entered in a public registry at the beginning of enrollment in order to be considered for publication. The ICMJE did not specify a registry, but provided criteria for a qualifying registry.

The ICMJE defines a clinical trial as any research project that prospectively assigns people or a group of people to an intervention, with or without concurrent comparison or control groups, to study the relationship between a health-related intervention *and* a health outcome. Health-related interventions are those used to modify a biomedical or health-related outcome; examples include drugs, surgical procedures, devices, behavioral treatments, educational programs, dietary interventions, quality improvement interventions, and process-of-care changes. Health outcomes are any biomedical or health-related measures obtained in patients or participants, including pharmacokinetic measures and adverse events. The ICMJE does not define the timing of first participant enrollment, but best practice dictates registration by the time of first participant consent [14, 15].

[ClinicalTrials.gov](http://www.clinicaltrials.gov) [16] and the International Standard Randomized Controlled Trial Number

Fact Box 50.6: Important Dates Related to ICMJE Data Sharing [14, 15]

- As of July 1, 2018, manuscripts submitted to ICMJE journals that report the results of clinical trials must contain a data-sharing statement.
- Clinical trials that begin enrolling participants on or after January 1, 2019, must include a data-sharing plan in the trial's registration. If the data-sharing plan changes after registration, this should be reflected in the statement submitted and published with the manuscript and updated in the registry record.

(ISRCTN) [17] meet those criteria and have become the most widely used registries to meet the ICMJE requirement.

These registries contain information about the clinical studies such as the investigation product under research, study objectives, study design, participating investigators and institutions, and funding.

The purpose of registering a clinical trial is to prevent bias and selective reported outcomes, to prevent unnecessary duplication of research, to keep the public informed about planned or ongoing trials, and to give ethics boards a platform to review similar works they may be considering.

The ICMJE now requires that authors include a plan for data sharing as a component of clinical trial registration. This plan must include where the researchers will house the data and, if not in a public repository, the mechanism by which they will provide others access to the data, as well as other data-sharing plan elements outlined in the 2015 Institute of Medicine report (e.g., whether data will be freely available to anyone upon request or only after application to and approval by a learned intermediary, whether a data use agreement will be required) [19, 20]. [ClinicalTrials.gov](http://www.clinicaltrials.gov) has provided data-sharing plans.

Data-sharing statements must indicate the following: whether individual de-identified participant data (including data dictionaries) will be

shared; what data in particular will be shared; whether additional, related documents will be available (e.g., study protocol, statistical analysis plan, etc.); when the data will become available and for how long; and by what access criteria data will be shared (including with whom, for what types of analyses, and by what mechanism).

As the clinical research industry moves toward increased transparency, the new researcher should be aware of the considerations when registering clinical trial results and the parameters surrounding manuscript publication.

50.4 Basic Principles, Considerations, and Essential Requirements of Any Clinical/Human Research Project

Prior to the start of any human clinical research, it is important to define the research and hypotheses, understand the experimental vs. investigational nature of the research, understand and be able to articulate risks, and to have a plan for risk mitigation to minimize risks prior to and during the study. Additionally, there are local, regional, and federal requirements for conduction of research involving human subjects that one must be aware of. Most institutions have criteria and other requirements for the evaluation, approval, and implementation of human research done on-site or with affiliates, so these are also important to understand prior to embarking on the human research project. Key responsibilities of the investigator, the institution, and those involved in supporting the study from a sponsorship perspective that are applicable to most human research projects are discussed below.

50.4.1 Responsibilities of the Investigator, Institution, and Sponsor

One of the most important responsibilities of each party involved is to ensure that the study is conducted in accordance with ethical principles

and related institutional and government requirements for the protection of the human subjects involved in the research.

The ethics of clinical research have developed from historical lessons that had few rules and regulations to protect the clinical subject. Some ethical guides that have influenced current ethics principles include the Nuremberg Code (1947), Declaration of Helsinki (2000), and the Belmont Report (1979). The National Institutes of Health [21, 22] defines seven main principles as a guide to conduct clinical research. We review these below, along with examples of questions that the researcher should consider in the development of an ethical and robust clinical program:

1. Social and clinical value: what is the potential value of this research to the subjects involved, to the community, and to the future of medicine and improvement in medical care?
2. Scientific validity: what study design features will reduce bias, and support appropriate hypotheses testing for the results to be interpretable, and applied to future research and/or medical practice?
3. Fair subject selection: what a priori criteria should be in place to avoid inclusion of subjects that are outside the specific population for research and to avoid exclusion of subjects due to bias? What screening practices and documentation of the rationale for inclusion and exclusion of subjects need to be in place to reduce potential for bias and to allow for appropriate subject selection into the study?
4. Favorable risk-benefit rationale: what clinical and protocol risks are evident? How are risks mitigated and managed prior to the study and during the study? How will subjects be informed of the risks prior to enrollment? What action will be taken if new risks are identified?
5. Independent review: what processes are in place to support the requirement for independent review of the protocol and informed consent prior to starting the study and during the study? Will a separate data review committee be provided to support safety and medical monitoring? Will the study be audited by an

- independent party or parties? Will FDA review be required prior to initiating the clinical research?
6. Informed consent (written and signed): what do subjects need to know about the research, the materials used in the research (e.g., drugs, biologics, devices, measures and tools applied, surgery or other interventions), the time commitment involved, their rights as a human subject, how their data will be used and what security will be applied to ensure privacy, and what to do if they have a question or an issue during the study, for example?
 7. Respect for potential and enrolled subjects: what effort will be made to ensure that subjects are properly treated with respect and dignity throughout the study? How will access to subject information be controlled, have limited access, and maintained? What training will study staff be given to ensure respect, privacy, and best research practices?

Two of the most important aspects of clinical research include (1) independent review committees to assess and review the protocol prior to approving clinical research and (2) the informed consent documentation, process, and assignment/training of personnel involved in administration of informed consent. Independent review committees (e.g., often referred to as an IRB or ethics committee) are comprised of a group of individuals who will assess the clinical trial and will review the protocol and any materials that are intended for the subject such as the informed consent, recruitment phone scripts, and advertisements for supporting community awareness of the study. The clinical trial should not begin until written approval of the protocol of the submitted documents is received. Only the approved documents may be used for the study, and if there are any changes desired to the documents approved, the committee must review and approve the changes prior to the use of those new materials. This oversight during human research supports proper consideration and independent perspective and reduces bias and the potential for gaps in human subject protection.

Each country as has a federal authority, such as the Food and Drug Administration in the United States, which oversees and allows the research to occur in accordance with federal law and regulation. In some cases, the federal authority may require review and approval of the research prior to initiation, in addition to the local/regional IRB or ethics committee review and approval. Most often the deciding factor as to pre-approval of research by a federal health authority is the risk to human subjects associated with the protocol and the risk associated with the materials or products involved. See Sect. 50.3, below, for additional details regarding federal health authority considerations and investigational medicinal products, therapies, and/or medical devices. The investigator should check with each country-specific regulatory authority for application requirements and verify the level of federal authority involvement with the IRB or ethics committee involved in research review.

50.4.2 Purpose of the Research

The purpose of the research and intent for data use are important to define early on. An investigator may be conducting clinical research for the purposes of publication, institutional quality improvement, addressing questions regarding the safety or efficacy of a new procedure, and/or assessing the safety and effectiveness of an investigational product. Plans as to what level of oversight and documentation required will be driven by the type of research involved, the purpose of the research, and what parties are interested.

Additionally, a clinical trial proposal should be supported by publications in a literature review that will serve as the foundation for manuscript references and also to support the proposed research scientific validity and study design. Be descriptive of the approach, discuss what was changed since previous works were conducted, and propose the new research to fully develop hypotheses. Pilot studies may be used to support hypothesis(es) and to demonstrate that subsequent studies are warranted.

50.4.3 Trial Operations Risk Management and Documentation to Support a Well-Controlled Trial

Several clinical trial documents, including essential documents [23, 24], listed should be in place prior to enrolling subjects in a clinical trial. Documents include, but are not limited to, the investigational plan, standard operating procedures, accountability of the investigational product, monitoring plan, data management plan, and statistical analysis plan. Deficiencies found during regulatory audit include inadequate monitoring, investigator noncompliance, informed consent deficiencies, and investigational product accountability. It is important to conduct risk assessment and provide a structured approach to risk mitigation. Evaluating the study sites, monitoring the frequency, and establishing a clinical events committee and/or data safety monitoring board are key to addressing operational deficiencies. For issues found in the protocol design, proactive measures should be taken so that they are not repeated in future studies. Also, swift correction of issues is important, along with documentation of the correction and follow-up to ensure that the correction made was effective action taken.

50.4.4 Subject Safety

Ethical conduct of a clinical trial includes protecting the rights, interests, and welfare of subjects throughout the duration of a clinical trial. Many study team members, such as the ethics committee, monitoring boards, and sponsor and investigator, play a role in subject safety. It is the investigator's responsibility to ensure ethics committee study approval(s) are in place, ethics committee continuing review, adherence to the protocol, the informed consent process is followed and the informed consent forms are signed and dated accordingly, adherence to GCP, adverse events are properly reported, and control of investigation product and to allow adequate time for monitoring the study. A full list of investigator

Fact Box 50.7: ICH GCP E6 Essential Documents [23, 24]

Essential Documents are those documents which are used to evaluate the quality of a clinical trial and the quality of the data produced. The various documents (abbreviated list) are grouped into three sections according to the stage of the trial during which they will normally be generated:

1. Before the clinical phase of the trial commences
 - (a) Investigator's Brochure (IB)
 - (b) Protocol
 - (c) Informed consent
 - (d) Case report forms (CRF)
 - (e) Agreements with investigator/institution, sponsor, and CRO
 - (f) IRB/EC/CA approvals
 - (g) Lab/test procedures
2. During the clinical conduct of the trial
 - (a) Updates/revisions to IB, protocol
 - (b) Approvals by EC/CA
 - (c) ICs and CRFs
 - (d) Adverse events
 - (e) Subject enrollment
3. After completion or termination of the trial
 - (a) Investigational product accountability
 - (b) Subject list
 - (c) Trial closeout
 - (d) Final clinical study reports

responsibilities can be found in the ICH Harmonized Guideline, Integrated Addendum to ICH E6 (R1): Guideline for Good Clinical Practice E6 (R2) [25, 26].

50.4.5 Detailed Protocol

A clinical trial protocol provides the background for the research and the design of the study and is the instructional tool for the investigational site(s) to conduct the study. ICH GCP [25, 26] provides guidance for the contents of a clinical protocol

and protocol amendments. Sections of a protocol should include:

- General Information
- Background Information
- Trial Objectives and Purpose
- Trial Design
- Selection and Withdrawal of Subjects
- Treatment of Subjects
- Assessment of Efficacy
- Assessment of Safety
- Statistics
- Direct Access to Source Data/Documents
- Quality Control and Quality Assurance
- Ethics
- Data Handling and Record Keeping
- Financing and Insurance
- Publication Policy
- Supplements

50.4.6 Qualified Medical Personnel and Trained Team

The principal investigator is responsible for the clinical trial conducted at his/her site. He/she may delegate qualified personnel to conduct certain aspects of the clinical trial; however the principal investigator ultimately remains responsible for their actions. A Delegation of Authority log is provided to each site, and each member of the clinical trial team will outline the duties they will perform. ICH GCP [25, 26] outlines the investigator's qualifications and agreements which outlines that the investigator should be qualified by education, training, and experience, they should be familiar with the investigational product, they should comply with GCP, the investigator/institution should permit monitoring and auditing, and the investigator should maintain a list of qualified persons to whom he/she has delegated trial-related duties.

50.4.7 Informed Consent Process

Properly conducting the informed consent process with potential subjects is extremely important to ensure subject protection. The process of

informed consent goes beyond a signature. The informed consent should allow for dialogue about the potential subject's participation. The potential subject should be given adequate time to review the information, consider all options, and ensure the potential subject understands what is being asked in order to participate and voluntarily agree to participate in the study. ICH GCP [25, 26] outlines informed consent of trial subjects in detail.

50.4.8 Data Collection and Documentation

Data collection can be paper-based or collected via electronic data capture (EDC), each having their pros and cons. No matter whether paper or electronic data capture is the method to be used in a clinical trial, the same supporting framework should be established to ensure minimal data errors. When developing case report forms (CRFs), ensure you are collecting pertinent data to be analyzed. Extra data points that are not meaningful to the research cause burden to the investigational sites. A sponsor or investigational site may use source document worksheets to capture data that they may not typically capture in private practice. Check with your regulatory agency to ensure worksheets are acceptable. Any type of source where the data is first documented is auditable, such as electronic medical records (EMR), radiology, and even the miscellaneous paper used to record data such as height, weight, and vital signs.

Ensure a data management plan (DMP) is in place prior to the study. Proper data quality management will help to ensure minimal errors and quality data that will support the study and hypothesis.

50.4.9 Safety Reports

The clinical protocol will define what constitutes an adverse event and serious adverse event (additional and differing definitions should be provided based on whether the investigational

Fact Box 50.8: ICH GCP Safety Reporting [27, 28]

All serious adverse events (SAEs) should be reported immediately to the sponsor except for those SAEs that the protocol or other documents (e.g., Investigator's Brochure) identify as not needing immediate reporting. The immediate reports should be followed promptly by detailed, written reports. The immediate and follow-up reports should identify subjects by unique code numbers assigned to the trial subjects rather than by the subjects' names, personal identification numbers, and/or addresses. The investigator should also comply with the applicable regulatory requirement(s) related to the reporting of unexpected serious adverse drug reactions to the regulatory authority(ies) and the IRB/IEC.

Adverse events and/or laboratory abnormalities identified in the protocol as critical to safety evaluations should be reported to the sponsor according to the reporting requirements and within the time periods specified by the sponsor in the protocol.

For reported deaths, the investigator should supply the sponsor and the IRB/IEC with any additional requested information (e.g., autopsy reports and terminal medical reports).

product is a drug or device). Regulations for reporting safety events vary from country to country, but they have similar procedures for classifying adverse events.

50.4.10 Intellectual Property and Inventions

The Clinical Trial Agreement (CTA) is the legally binding document between the investigator and the sponsor. Sections include, but are not limited to, publication, data ownership, intellectual property, indemnification and insurance, adverse event reimbursement, and the study budget.

The definition of intellectual property will vary between each country's patent laws. The scope of an invention and ownership of the data will be defined and agreed to in the CTA.

50.5 Regulatory Considerations: Competent Authorities and Research Leading Toward a Commercial Product

Clinical research has as its goals to protect the patient or subject but also to contribute or generate knowledge which can be of use to patients or other clinicians or researchers. Regulatory requirements for clinical research studies vary depending on the type of investigation. If the new researcher is involved in any research which require human clinical subjects or could lead to a commercial product, they need to be aware of some additional considerations so that the research effort can be utilized to its fullest potential. The new researcher should begin with an understanding of the regulatory considerations and differences of the commercial market(s) they may target for a regulatory submission. The new researcher should familiarize themselves between device, drug, and biologic regulations as they pertain to the product and to the jurisdiction where regulatory approval may be sought. For many new drugs and biologics, clinical studies will be required to demonstrate safety and effectiveness. For devices in many markets, the landscape can become even more complex as they can be assigned to one of three regulatory classes (Class I, Class II, and Class III) based on the level of risk or control necessary to assure the safety and effectiveness of the device. This device classification then defines the regulatory requirements for an approval of a new device and regulatory control, and requirements increase from Class I to Class II to Class III. Depending on the device classification, clinical studies may or may not be required.

Each country has its own regulatory authority with its own regulations for approving clinical study protocols and also for conducting clinical studies when testing and approving a new device, drug, or biologic. The following table lists the

major global regulatory or competent authorities (Tables 50.1, 50.2, and 50.3).

The websites can provide a source of information regarding regulatory requirements, processes, or expectations with regard to all aspects of the product. The competent authorities may also offer guidance documents which more definitively describe information and data requirements for the development, clinical studies, and regulatory process for defined indications. The guidance documents may also outline the expectations for each phase of clinical development (Phases I–III).

The competent authorities, singular or multiple in some cases, oversee clinical studies in their respective countries. Initially their responsibilities include reviewing and approving clinical trial protocols and then ensuring that clinical trials comply with national regulations and international guidelines. After the clinical studies, development, and subsequent approval, they also have quality assurance authority to ensure the production, distribution, labeling, and safety monitoring of new or existing devices, drugs, and biologics.

Table 50.1 National competent authorities in major global markets: European Medicines Agency

European Medicines Agency		
Austria	Austrian Agency for Health and Food Safety	http://www.ages.at/
Belgium	Federal Agency for Medicines and Health Products	www.fagg-afmps.be/
Croatia	Agency for Medicinal Products and Medical Devices (HALMED)	http://www.halmed.hr/en/
Denmark	Danish Health and Medicines Authority	www.laegemiddelstyrelsen.dk
Germany	Federal Office of Consumer Protection and Food Safety	www.bvl.bund.de
Greece	National Organization for Medicines	www.eof.gr
Italy	Ministry of Health	http://www.salute.gov.it/
Netherlands	Healthcare Inspectorate	www.igz.nl
Poland	Office for Registration of Medicinal Products, Medical Devices, and Biocidal Products	www.bip.urpl.gov.pl
Spain	Spanish Agency for Medicines and Health Products	www.aemps.gob.es
United Kingdom	Medicines and Healthcare products Regulatory Agency	https://www.gov.uk/government/organisations/medicines-and-healthcare-products-regulatory-agency

Table 50.2 Health authorities in Latin America

Health authorities in Latin America		
Brazil	Ministério da Saúde	http://portalms.saude.gov.br/
Mexico	Secretaría de Salud	https://www.gob.mx/salud
Argentina	Ministerio de Salud	https://www.argentina.gob.ar/salud

Table 50.3 Health authorities in Latin America

Health authorities in Asia-Pacific		
India	Central Drugs Standard Control Organization	https://cdscoonline.gov.in/CDSCO/homepage
China	China Food and Drug Administration	http://eng.sfda.gov.cn/WS03/CL0755/
Japan	Ministry of Health, Labour and Welfare	http://www.mhlw.go.jp/english/index.html
Australia	Therapeutic Goods Administration	www.tga.gov.au
Health authorities in North America		
Canada	Health Canada	https://www.canada.ca/en/health-canada.html
United States	Food and Drug Administration	www.fda.gov

50.5.1 Role of Sponsors and Clinical Research Organizations (CRO)

If the research involves a sponsor, many times the investigator has limited interaction with the competent authority. A clinical trial sponsor can be an individual, an investigator, company, institution, or organization that takes responsibility for the initiation, management, and financing of a clinical trial. A sponsor can be a device, pharmaceutical or biotech company, a non-profit organization such as a research fund, a government organization or the institution where the trial is to be conducted, or the individual investigator. The sponsor will assume the responsibilities such as protocol development, financing the trial, and will seek permission for trial initiation from the competent authority or authorities. The competent authority will interact with the sponsor and approves the trial protocol that is provided to the investigator. The sponsor can have the following responsibilities:

- Submitting the plan for the clinical trial to the competent authority for approval
- Informing clinical investigators about the test article, its safety and instructions for use, as well as training the staff and facility regarding its use and handling
- Making sure there are an appropriate number of test articles for the investigation
- Ensuring the trial protocol is properly reviewed by an experienced EC
- Monitoring the trial to ensure the protocol is being followed, data collection is accurate, adverse events are reviewed and reported, and all regulations are complied with

If the research involves a clinical research organization (CRO), many of the responsibilities of the investigator are delegated to the CRO. CROs are independent companies providing research services for the device, drug, and biologic industry and function as outsourcing of tasks related to clinical trials. Such outsourcing services can be related to project management, trial monitoring, data collection, and medical statistics work. When a CRO is contracted by a sponsor, it and their designated clinical trial monitor (CTM) or clinical research associate (CRA) takes on many and

sometimes all of the sponsor's trial responsibilities. In some cases, central laboratory services are an important ingredient of clinical trials, conducting work such as processing blood samples and reading radiographic images. Sponsors and sometimes competent authorities may require that one single-source process sample so there is a standardized process, results are reliable and reproducible, and data is collected and stored in a centralized facility. The central laboratory services can be considered a specialized CRO.

Of note, investigators must be able to report all results of the clinical trial, regardless of outcome. In the United States, there is a requirement that results of trials of FDA-approved products to be posted within 12 months of trial completion on www.clinicaltrials.gov.

50.6 Elements of the Technical Document

At the conclusion of the clinical research on a device, drug, or biologic, the next steps may involve seeking regulatory approval for commercial marketing of the investigational product. The compilation of the technical file is a critical step in the regulatory approval process and includes detailed information about the design, function, composition, use, claims, and clinical evaluation of your drug, device, or biologic. The Common Technical Document (CTD) was designed to provide a common format between Europe, the United States, and Japan for the technical documentation included in an application for the registration of a human device, drug, or biologic product. The agreement to harmonize and assemble all the quality, safety, and efficacy information in a common format was a major advance in the global regulatory review process and enabled implementation of good review practices. For device, drug, or biologic companies, it eliminated the need to reformat the information for submission to the different ICH competent authorities. A CTD is a comprehensive description of your product and demonstrates compliance with the requirements of the applicable regulatory requirements. For devices in the European Union, the directives that need to be met include Medical Devices

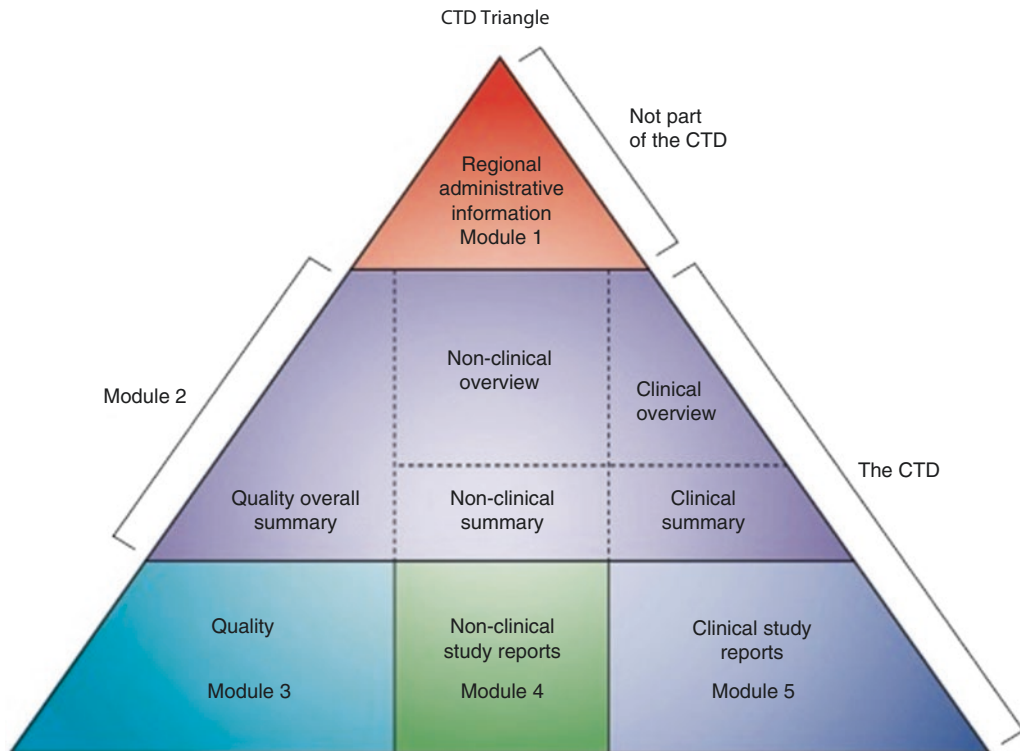
Directive 93/42/EEC (MDD) (transitioning to Medical Device Regulation 2017/745 (MDR)), In Vitro Diagnostic Medical Devices Directive 98/79/EC (IVDD), and Active Implantable Medical Devices Directive 90/385/EEC (AIMDD).

ER#	Scope
6a	Clinical evaluation
<i>Part II—design and construction requirements</i>	
7	Chemical, physical and biological properties
8	Infection and microbial contamination
9	Construction and environment properties
10	Device with a measuring function
11	Protection against radiation
12	Devices with an energy source
13	Information supplied by the manufacturer

50.7 Essential Requirements for Medical Devices: MDD Annex I, 93/42/EEC

ER#	Scope
<i>Part I—general requirements</i>	
1	Risk reduction and acceptable risk/benefit
2	Safety and risk controls
3	Intended performances
4	Lifetime of the devices
5	Transportation and storage
6	Side-effects must constitute acceptable risk

The CTD is organized into five modules [29, 30] (Fig. 50.1). Module 1 is region specific, and Modules 2, 3, 4, and 5 are intended to be common for all regions. Please refer to the ICH guidances for industry: M4Q The CTD, Quality [31, 32]; M4S The CTD, Safety [33, 34]; and M4E The CTD, Efficacy [35] for additional information [36].



The CTD triangle. The common Technical Document is organized into five modules. Module 1 is region specific and modules 2, 3, 4 and 5 are intended to be common for all regions.

Fig. 50.1 The Common Technical Document triangle has five modules. Module 1 is region specific; Modules 2–5 are common to all regions [30]

Module 1: Administrative Information and Prescribing Information

According to the FDA guidance document, “This module should contain documents specific to each region; for example, application forms or the proposed label for use in the region. The content and format of this module can be specified by the relevant regulatory authorities.” The content of this module is specific to each market or competent authority and may include the following information:

- Introductory letter
- Application form
- Product information
- Labeling and Information for use
- Product information for products already approved
- Information about the experts
- Country-specific requirements
- Environmental risks
- Others

Module 2: Overviews and Summaries of Modules 3–5

This module summarizes and outlines the information that will be presented in Modules 3–5. It should begin with a general introduction to the device, drug, or biologic, including its class, mode of action, and proposed clinical use. In general, this introduction should be a summary and limited in overall length:

- Table of Contents
- Introduction
- Quality Overall Summary (QOS)
- Nonclinical Overview
- Clinical Overview
- Nonclinical Written and Tabulated Summaries, including pharmacology, pharmacokinetics, toxicology, biocompatibility, viral inactivation
- Clinical Summary, including clinical pharmacology studies, clinical safety and efficacy, literature references, and synopses of individual studies

Module 3: Quality (Device, Drug, or Biologic Documentation)

This module provides information on the quality of the product and should be presented in the structured format described in the ICH M4Q guidance. This module contains the detailed chemical, pharmaceutical, and biological data relevant to the product. This module will consist of items such as:

- Table of Contents
- Body of Data
 - Product—information and its properties
 - Manufacturing—details, process, control of materials, validation
 - Characterization
 - Controls
 - Reference standards
 - Packaging
 - Stability
 - Appendices—facilities, regional information requirements
- References

Module 4: Nonclinical Reports (Pharmacology/Toxicology/Biocompatibility)

This module presents the integrated and critical information on the pharmacologic, pharmacokinetic, and toxicological evaluation of the drug, device, or biologic. This module consists of the reports which were summarized in Module 2, so all the complete and detailed reports will be provided in this module. The nonclinical study reports should be presented in the order described in the ICH M4S guidance. This module will consist of items such as:

- Table of Contents
- Pharmacology
- Pharmacokinetics
- Toxicology
- Literature references

Module 5: Clinical Study Reports (Clinical Trials)

This section is intended to provide the full complete and detailed reports from all of the available clinical information. This includes clinical study reports, information obtained from any meta-analyses or other cross-study analyses, and post-marketing data for products that have been marketed in other regions. This module consists of the reports which were summarized in Module 2. The human study reports and related information should be presented in the order described in the ICH M4E guidance. This module will consist of items such as:

- Table of Contents
- Tabular Listings of All Clinical Studies
- Clinical Study Reports—safety, efficacy, dosing, post-marketing, case reports
- Literature references

Clinical Vignettes

One area that has been the subject of extensive regulation has been in the use of cell therapies to treat musculoskeletal injuries. In the United States for over 20 years, the only cell therapy treatment approved for use for articular cartilage injuries was the first-generation autologous chondrocyte implantation technique. The regulatory environment in the United States limited the use and evaluation of other cell-based therapies and contributed to a relative lack of progress in the United States regarding cell therapy treatment compared to other countries. In this situation, addressing the regulations around approval of new therapeutic cell therapies has taken an extensive amount of time. Protocol design and development involved multiple meetings with the FDA over a

period of years. Regulatory guidance from the FDA also changed during this time, resulting in further changes to the protocol. Ultimately, several protocols for more advanced cell-scaffold-based autologous chondrocyte implantation techniques were approved, and are currently in various stages of the trial process, with one technique recently approved for clinical use.

50.8 Summary

Regulation is an inescapable aspect of any research activity that involves human or animal subjects. The reasons behind this regulation are complex and can involve a number of issues including human and animal subject safety, patient protection, and social, ethical, and political considerations. Regulation can result in increased complexity in study design and execution. It is critical for investigators to be familiar with the regulatory issues involved in research and to be careful to comply with required regulation. Failure to comply with regulation can result in subject risk or harm and for the investigator can result in possible civil liability or criminal charges.

Take-Home Messages

- Regulatory issues govern the conduct of research and are typically based in protecting the health and safety of subjects, but can also be related to ethical issues, such as the ban on aspects of human cloning due to “human dignity” concerns.
- The regulatory burden often depends on the perceived risks; new drugs and devices typically undergo an expensive and time-consuming multistage approval process designed to decrease risks, whereas drugs or devices similar to predicates may be seen as posing lower risk and often receive expedited approval.
- Local regulations often exist, but there is increasing harmonization internationally

around Good Clinical Practice (quality, safety, efficacy, and multidisciplinary) guidance, including documentation.

- Research should be performed in accordance with ethical principles.
- Failure to comply with regulation can result in subject harm and can result in civil liability or criminal charges.

References

1. ICH GCP. <http://ichgcp.net/>. Accessed 18 July 2018.
2. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R1_Guideline.pdf
3. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Guideline for good clinical practice E6(R1); 1996. p. 59.
4. http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002874.pdf
5. Committee for Human Medicinal Products. Guideline for good clinical practice E6(R2). London: European Medicines Agency; 2017.
6. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R2_Step_4_Presentation_06Feb2017.pdf.
7. ICH E6(R2) Expert Working Group. Integrated addendum to ICH E6(R1): guideline for good clinical practice E6(R2); 2017.
8. <https://about.citiprogram.org/en/series/good-clinical-practice-gcp/>.
9. Collaborative Institutional Training Initiative (CITI Program). Good clinical practice (GCP)—CITI Program. <https://about.citiprogram.org/en/series/good-clinical-practice-gcp/>. Accessed 18 July 2018.
10. http://www.barnettinternational.com/Publications/Good-Clinical-Practice%2D%2DA-Question%2D%2D-Answer-Reference-Guide-2017/?gclid=EAIAIqobChMI783lydi_2QIVyLbACh0ORwRVEAAYASAAEgL82vD_BwE.
11. Good clinical practice: a question & answer reference guide 2018 (electronic). Needham: Barnett International Publication; 2018. <https://www.barnettinternational.com/good-clinical-practice-a-question-amp-answer-reference-guide-2018-electronic-0>.
12. <http://www.clinicaldevice.com/mall/Workshops.aspx>.
13. Medical Device Workshops. <http://www.clinicaldevice.com/mall/Workshops.aspx>. Accessed 18 July 2018.
14. <http://icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html>.
15. International Committee of Medical Journal Editors. ICMJE | recommendations | clinical trials. <http://icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html>. Accessed July 18, 2018.
16. National Library of Medicine. <https://www.clinicaltrials.gov/>. Accessed 18 July 2018.
17. ISRCTN Registry. <http://www.isrctn.com/>. Accessed 18 July 2018.
18. World Health Organization. International Clinical Trials Registry Platform (ICTRP). <http://www.who.int/ictrp/network/primary/en/>. Accessed 18 July 2018.
19. <http://www.nejm.org/doi/full/10.1056/NEJMe1515172>.
20. Taichman DB, Backus J, Baethge C, Bauchner H, de Leeuw PW, Drazen JM, et al. Sharing clinical trial data—a proposal from the International Committee of Medical Journal Editors. *N Engl J Med*. 2016;374:384–6. <https://doi.org/10.1056/NEJMe1515172>.
21. <https://clinicalcenter.nih.gov/recruit/ethics.html>.
22. National Institutes of Health. NIH Clinical Center: Ethics in Clinical Research. <https://www.ncbi.nlm.nih.gov/pubmed/>. Accessed 18 July 2018.
23. <http://ichgcp.net/8-essential-documents-for-the-conduct-of-a-clinical-trial>.
24. Essential documents for the conduct of a clinical trial. Documents. <http://ichgcp.net/8-essential-documents-for-the-conduct-of-a-clinical-trial>. Accessed 18 July 2018.
25. http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E6/E6_R2_Step_4_2016_1109.pdf.
26. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). Integrated addendum TO ICH E6(R1): guideline for good clinical practice E6(R2). ICH; 2016.
27. <http://ichgcp.net/411-safety-reporting>.
28. Safety Reporting. <http://ichgcp.net/411-safety-reporting>. Accessed 18 July 2018.
29. <http://www.ich.org/products/ctd.html>.
30. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). CTD: ICH: ICH; <http://www.ich.org/products/ctd.html>. Accessed 18 July 2018.
31. <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM073280.pdf>.
32. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research. Guidance for industry M4Q: the CTD—quality. Rockville: U.S. Department of Health and Human Services; 2001.
33. <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM073299.pdf>.
34. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research. Guidance for industry M4S: The CTD—safety. Rockville: U.S. Department of Health and Human Services; 2001.

-
35. US Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research. Guidance for industry M4E: The CTD—efficacy. Rockville: U.S. Department of Health and Human Services; 2001.
 36. <https://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM073290.pdf>.



What Is Needed to Make Collaboration Work?

51

Richard E. Debski and Gerald A. Ferrer

51.1 Introduction: Importance of Collaborations

Due to the complexity of many healthcare problems, finding the appropriate solutions may not be feasible without the right partners. Collaborations between basic scientists and clinicians are critical to advancing scientific knowledge by performing high-quality, high-impact, translational research [1–4]. Through collaboration, much more can be achieved. Collaboration offers numerous benefits to both parties in addition to completing research studies. On an individual level, collaboration offers the opportunity to expand one’s knowledge and skill set while fostering a professional network that could develop into lifelong relationships. Professionally, collaboration allows for access to new resources and increased prestige by publishing in more impactful journals, which increases the likelihood of having their work implemented to clinical practice. Achieving a successful collaboration between basic scientists and clinicians is not a simple task. This chapter will discuss key components needed to make collaborations between basic scientists and clinicians successful.

R. E. Debski (✉) · G. A. Ferrer
Orthopaedic Robotics Laboratory, Department of
Bioengineering, Swanson School of Engineering,
University of Pittsburgh, Pittsburgh, PA, USA
e-mail: genesis1@pitt.edu

51.2 Leadership

Effective leadership is essential to the success of any collaboration. Effective leaders need to possess strong communication and management skills. Leaders must be able to communicate the overall vision of the research group, as well as the roles and responsibilities of each team member to accomplish the group’s goals. Without a vision, team members may be working without a purpose of what they are doing and why. One good way to assess if each team member understands the vision of the research group is have them give a “lab tour.” Lab tours offer the opportunity to explain different projects conducted in the lab and their contribution to fulfilling the research mission of the lab.

Many research teams have a diverse group of individuals with different levels of experience and expertise. From undergraduate students to young investigators or international research fellows, every team member should know his/her responsibilities and be held accountable by the leaders of the group. While leaders should keep every team member accountable, effective leaders are able to accomplish this without promoting an aggressive environment. One important factor that may lead to an unhealthy work environment is not giving team members the proper credit or recognition. For example, the leaders need to communicate the criteria for being an author on abstracts and publications, authorship order, and

Fact Box 51.1

- Communication and management skills are keys toward effective leadership.
- Have a shared vision of scientific goals.
- Give appropriate credit and recognition.

who is responsible for presenting at conferences. At our lab, every author must fulfill at least three of the five requirements:

1. Writing and providing feedback
2. Developing and performing experimental methods
3. Data processing
4. Project idea generation
5. Acquisition of project funding

Receiving appropriate credit and recognition is particularly important for aspiring young team members who are trying to promote their careers. By fostering a positive work environment and defining the roles of the team members, achieving goals that work toward fulfilling the mission of the research team becomes much easier. Ultimately, the leaders of the collaboration should focus on developing a supportive and fair environment where the group can conduct high-quality research.

51.3 Building Your Team

When pursuing a potential collaboration, it is important to be selective and get to know the person who you want to collaborate with. One must assess whether they are the “right” collaborator who they trust and envision having a lasting relationship to accomplish their scientific goals at a high level of research quality and productivity. Finding the “right” collaborator may take time and require a period of trial and error. Potential collaborations can often be found from mutual acquaintances or one-on-one discussions at scientific conferences. While basic scientists focus on performing research, the clinicians’ primary focus may be on their patients’ care and other clinical responsibilities. Thus, it is imperative for

the basic scientist to consider and assess the clinician’s commitment and dedicated time for research. Some clinicians may be very involved throughout the entire process of a research project, from idea generation, developing methods, data analysis, and writing. There are clinicians who may provide occasional support during the research process, and others may be interested in generating project ideas and reviewing the final abstract or manuscript. Given the significant responsibilities of clinicians, it is important to gauge how much time and effort each party is willing to contribute to the collaboration. For example, clinicians may only be able to meet on the weekends or very early in the morning or late at night. Additionally, meetings may be most appropriate in the lab or office space of the scientist. Both parties need to be willing to compromise and make sacrifices for a collaboration to work and be long lasting.

Another important factor to consider when building an effective research team is trust (Fig. 51.1). Trust entails understanding the other team member’s ambitions, desires, and values. Knowledge of what drives each team member may help recruit the right personnel for what your group may want to accomplish. Furthermore,

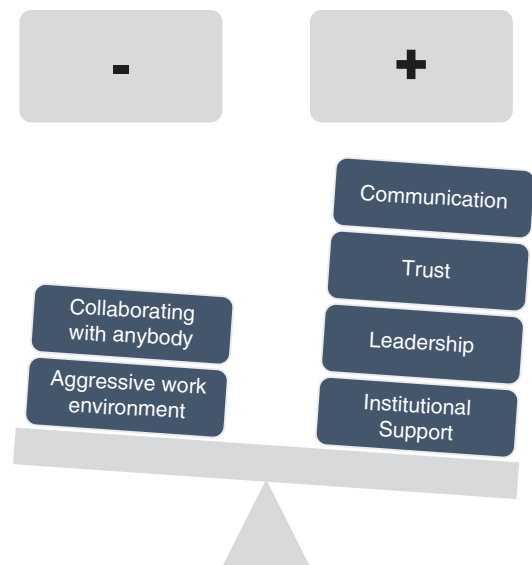


Fig. 51.1 Key building blocks needed to develop a positive and successful collaboration

Fact Box 51.2

- Choose the “right” collaborator, not just anyone.
- Long-lasting relationship
- Develop trust within team.

Fact Box 51.3

- Consistent terminology and limited technical jargon
- Promote disagreement while containing conflicts

trust within the team is predicated on reliability. To be an effective team, everyone needs to be relied upon to do their job, such as meeting deadlines and fulfilling expectations. In addition, team members must trust each other’s abilities and expertise. There needs to be a balance of keeping each other accountable for their work without wasting time and resources. As a team member, it is important to trust one another and be trustworthy yourself, in terms of your character and work. Without trust, the productivity of the research team and quality of work may be compromised.

important to establish. At our lab, we find our weekly lab meetings a perfect opportunity for diverse opinions. During the lab meetings, we critically evaluate each project being conducted within our research group and discuss lab philosophy. The lab meetings present an opportunity to not only critically evaluate the projects but also provide the presenter the opportunity to practice effective communication of their project to a diverse audience (i.e., undergraduate students, Ph.D. students, medical students, residents, fellows, clinicians, engineers, professors, etc.).

51.4 Communication

Communicating with people within one’s own discipline can be difficult. Interdisciplinary communication is even more challenging, yet essential for the success of collaboration between basic scientists and clinicians. Ineffective communication can be due to not understanding each other and is generally due to poor terminology and overuse of technical jargon. When communicating between disciplines, it is important to use consistent terminology to what is accepted by the scientific community and limit technical jargon. Thus, effective communication can be achieved within your own research group, as well as the scientific community at large. Furthermore, effective communication can improve research efforts to produce high-quality studies. The key is to promote a diversity of opinions between team members. Diverse opinions may promote disagreement, but they can push your collaboration to become better (i.e., strengthen relationships and trust, new ideas and solutions to problems). However, too much disagreement can lead to major conflicts that drive collaborations apart. A formal mechanism for conflict resolution is

51.5 Institutional Support

Support from your institution for research collaboration efforts cannot be understated. Institutional support that provides adequate time, funding, facilities, equipment, and personnel puts your collaboration efforts in a position to succeed. Moreover, institutional support can help maintain long-lasting collaborations and allow performance of research at a high level. For example, our lab was developed through the support of the Departments of Orthopaedic Surgery and Bioengineering. Their combined support provides an ideal environment for training and completion of research projects in a multidisciplinary environment.

51.6 Have Fun!

At the end of the day, what makes collaborations work is truly enjoying what you are doing and who you are doing it with. Working with people who share the same amount of enthusiasm and passion as you about what you are doing is fun. Answering complicated healthcare questions and pioneering scientific breakthroughs are very

rewarding. As such, it is important to celebrate accomplishments of not only the research group but of individuals as well.

Take-Home Message

- What makes collaborations between basic scientists and clinicians work is not an easy task and cannot be accomplished alone.
- First and foremost, it is important to find the right collaborator who you believe you can have a lasting relationship with and shares a common vision of research aspirations.
- To fulfill your vision, build a team that exemplifies three main points discussed throughout this chapter: (1) leadership, (2) trust, and (3) communication.
- In addition, institutional support helps maximize productivity and longevity of collaborations.

References

1. Bennett LM, Gadlin H. Collaboration and team science: from theory to practice. *J Investig Med.* 2012;60:768–75.
2. National Research Council. *Facilitating interdisciplinary research.* Washington, DC: The National Academies Press; 2004.
3. Green BN, Johnson CD. Interprofessional collaboration in research, education, and clinical practice: working together for a better future. *J Chiropr Educ.* 2015;29:1–10.
4. Mattessich PW, Monsey BR. *Collaboration: what makes it work. A review of research literature on factors influencing successful collaboration.* St. Paul: ERIC, Amherst H. Wilder Foundation; 1992.



Aleksei Dingel, Jayson Murray, James Carey,
Deborah Cummins, and Kevin Shea

52.1 Introduction

52.1.1 Why Clinical Practice Guidelines?

Within the past two decades, there has been a push toward evidence-based clinical practice guidelines (CPGs). These guidelines, unlike their opinion-based predecessors, would be designed to streamline health-care efficiency, improve health-care outcomes, and decrease practice variation [1]. In 2008, the US Congress mandated that the Institute of Medicine (IOM) develop standards for the evidence-based guidelines. In response, the IOM produced a rubric for well-organized and reproducible guideline development and evidence-based systematic review.

A. Dingel · K. Shea (✉)
Department of Orthopaedic Surgery, Stanford
University Medical Center, Stanford, CA, USA
e-mail: adingel@stanford.edu; kgshea@stanford.edu

J. Murray · D. Cummins
American Academy for Orthopaedic Surgeons
(AAOS), Norridge, IL, USA
e-mail: jmurray@aaos.org; cummins@aaos.org

J. Carey
Pennsylvania Hospital, Philadelphia, PA, USA
e-mail: james.carey@uphs.upenn.edu

52.1.2 A New System to Translate Best Evidence into Best Practice

Historically, clinical practice guidelines were largely based upon the consensus of physician expert opinion, specialist group recommendation, governments, payers, etc. [2]. Unfortunately, these unregulated recommendations frequently contradicted each other. The lack of a consistent and reproducible recommendation development process led to variations in patient care and questions about the validity and reliability of the guideline process. Other questions about the CPG process included concerns about the management of conflicts of interest (COI), as well as the ranking of relevant evidence [2]. Gaps in evidence, poor quality reviews, and biased recommendations based off of lower levels of research were all concerns relating to the guideline development [1]. Consensus/expert opinion-based guidelines left much to be desired by patients and caregivers alike and these concerns were sufficient to warrant a call to change from consensus-based to evidence-based guidelines.

Evidence-based clinical practice guidelines have been designed to replace the consensus-based guideline to increase health-care efficiency and patient care success.

In 2001, the IOM Committee on Quality of Health Care in America completed an extensive analysis of the health-care system and concluded that there were four key quality problem areas:

- The growing complexity of science and technology.
- The increase in chronic conditions.
- A poorly organized delivery system.
- Constraints on exploiting the revolution in information technology.

These quality problem areas, along with questions raised regarding trustworthy and appropriate development of consensus–/recommendation-based guidelines, lead the IOM, along with other health-care agencies, to call for the increased use of evidence-based clinical practice guidelines (CPGs) in order to improve quality of care, decrease inefficiencies, and reduce practice variation within the health-care system in America [2, 3].

Although financial benefits are not the main focus of an evidence-based practice guideline, improved guidelines may also reduce costs [4].

52.1.3 Quality Problem Areas

52.1.3.1 Science and Technology

The rapid advancement of science and technology in health care creates challenges for improving the safe, effective, and efficient delivery of health care [5]. In addition, boundless medical research databases have led to challenges for physicians, patients, and payers, who desire access to timely, concise, relevant information to guide care. The volume of relevant scientific information provided by the extensive literature databases is overwhelming, and a process to select the highest-quality, lowest-bias research from these databases is critical to practice evidence-based practice.

Over the past 30 years, the number of randomized clinical trials alone has increased from just over 100 to nearly 10,000 *annually*. The last 5 years alone account for nearly 50% of published articles in the medical literature, and there is no evidence the rate of publication is slowing [6].

There is no doubt the data organization and filter system of this information are in desperate need of a remodel. Rather than sorting through endless clinical trials to determine the best course of action, providers and patients should be able to turn to trustworthy evidence-based guidelines to efficiently determine the appropriate route of care [7].

52.1.3.2 Chronic Conditions

Chronic conditions, defined by the Centers for Disease Control and Prevention (CDC) as any illness lasting longer than 3 months and not self-limiting, were the leading cause of illness, disability, and death in America in 1996 [8]. According to a 2008 survey conducted by the CDC/National Center for Health Statistics, 85.6% of individuals 65 and older have *at least* one of the following chronic conditions: arthritis, asthma, cancer, cardiovascular disease, chronic obstructive pulmonary disease, and diabetes. In 2030, when the large baby boom cohort has entered old age, one in five persons is expected to be in this senior age group. With modern medicine and technological advances adding years to the average American life expectancy, now over 76 years of age, the incidence and prevalence of chronic conditions will only increase [3, 9]. According to the CDC, in 2012, almost half of all Americans (117 million people) were living with one or more chronic conditions [10], and in 2014, seven of the top ten leading causes of death were chronic diseases [10]. The treatment of chronic conditions accounted for 62% of health-care

spending in 2008 [10, 11], and in 2012, that number grew to 83%. Not to mention those with five or more chronic conditions had an average of almost 15 physician visits and filled over 50 prescriptions in a year. Osteoarthritis, a degenerative joint disease, affected 54 million Americans in 2014. According to the CDC, that number will rise to 67 million in the year 2025 [12].

The treatment of chronic conditions accounted for 62% of health-care spending in 2008 [10, 11], and in 2012, that number grew to 83%.

The demographic transformations that are projected to occur over the following years have important implications for the organization of the health-care delivery system. Self-management, family support services, committing to the treatment plan, and sustained follow-up visits are just as vital to patient recovery as initial diagnosis. Collaboration between the health-care provider, health-care provider team, patients, and patient's family adds an additional layer of complexity that must be considered when developing clinical guidelines [3]. It is yet another need for universally applied, clear, concise, and streamlined medical guidelines.

The demographic transformations that are projected to occur over the following years have important implications for the organization of the health-care delivery system.

52.1.3.3 Poorly Organized System

The current health-care delivery system is a labyrinth, a seemingly endless web of non-answers. Patients and families have described it as a “nightmare to navigate” [13]. Clinicians have reflected that it is an acute waste of time. The complex series of hand-offs between doctors, specialists, hospitals, insurance agencies, third parties, and other providers decreases the efficiency of patient care. While multiple hand-offs from specialty clinician to specialty clinician are vital when treat-

ing persons with multiple chronic conditions, the current mechanism of coordination is lacking and needs reconfiguring to increase efficiency and ensure safety and proper treatment. The ultimate goal is to help, not hinder a patient. It appears obvious that coordination should be as smooth and with the least number of hand-offs as possible to minimize time delay in health-care delivery.

52.1.3.4 Constraints on Information Technology

Information technology poses serious concerns for many health-care providers, the main concern being a patient's misunderstanding of proper medical treatment as they turn to web-based self-diagnosis and treatment rather than taking the time to see a trained medical professional. This may lead to serious illnesses being left poorly or inadequately treated.

However, when appropriately applied, the information technology is also a great tool to patients. E-mail allows for efficient communication between provider and patient. The web allows patients to self-educate and take more responsibility for and control of their recovery process. Online forums have been beneficial, especially for those struggling with rare diseases that may not have a community near them which they can lean on. Information technology has also the potential to increase the quality of health care through improving physician communication and removing communication barriers to health-care delivery.

These problems outlined by the IOM are simply additional reasons to update the health system and guideline development process.

52.2 What Is an Evidence-Based Clinical Practice Guideline?

CPGs, as defined by the CPG Development Manual, are *statements that include recommendations intended to optimize patient care that are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options* [14].

These evidence-based recommendations are developed using a minimally biased, transparent,

and reproducible evaluation of published medical literature. Evidence-based CPGs are designed to withstand the type of scrutiny and review its “expert group/consensus-based” predecessor could not. They serve as an effective synthesis of an enormous literature database, providing a complete yet concise summary of available knowledge and a detailed treatment plan for a specific topic or condition. These “evidence-based guidelines” undergo a rigorous protocol to deliver the optimum care route for the patient [2]. Such a guideline should streamline patient care while ensuring patient safety and increasing outcome success. As with all information and technologies, CPGs are subject to regular updates as new research and clinical studies are published [3]. CPGs are beneficial in that they provide an efficient source of information for the best course of treatment while allowing for flexibility in a treatment pathway [3].

52.2.1 Trustworthy CPGs

The need for *trustworthy* guidelines is one of the main driving factors for the new guidelines. Guidelines must be developed by a qualified and diverse group of individuals. The development process critically analyzes the data, the source of the data, and those who conducted the study to ensure limited guideline bias. Bias and COIs can influence the efficacy of and impact published research findings have on a community [15–17]. Bias must be minimized in order to provide the public with the most trustworthy guideline. When bias and COIs are allowed to traverse the boundary line between good research and bad, the effect can be detrimental.

For instance, therapeutic drug research often is run more like a promotional campaign for pharmaceutical companies, rather than a clinical research study, intended to increase sales rather than improve drug performance [18]. In 2008 the *New England Journal of Medicine* published a study [19] which reviewed the selective publication process of antidepressant drugs and the effect those selected publications have on drug efficacy. The study found that of the 74 clinical

trials conducted on one specific antidepressant, 38 produced positive results and 36 found the drug to have “questionable or no efficacy” [18]. However, only 8% of the “questionable or no efficacy” studies were published, while 94% of the positive studies were published. Moreover, 15% of the 8% “questionable or no efficacy” studies were published in such a way as to spin the results in a positive form [19]. As drug companies can cherry pick which data they wish to present, it is easy for physicians and medical providers to inadvertently develop a biased opinion about the drug. This can influence clinical practice and prescribing habits. Unsurprisingly, additional studies have found that industry-sponsored studies are significantly more likely to report favorable results and less likely to report unfavorable outcomes than their federally funded counterparts [15, 17]. This is troubling as many drugs are associated with serious adverse effects.

As such, it is vital that developers of CPGs look closely at the research evidence and develop CPGs in a trustworthy, reproducible, and transparent manner. Developers must consider not only the findings but also who sponsored the study. They must rigorously scrutinize if the results are reported truthfully. Only then can guideline development occur with minimal bias and maintain reliability.

52.3 Development of CPGs

In response to Congress and to develop trustworthy guidelines, the IOM has established eight standards of developing CPGs [1]:

1. Transparency.
2. Management of conflicts of interest (COI).
3. Development group diversity.
4. Systematic review.
5. Evidence and recommendation strength.
6. Articulation of recommendations.
7. External review.
8. Updating.

Each of these standards is intended to create the most well-researched, trustworthy, and clini-

cally relevant guideline possible. These standards, or “guidelines for the guidelines,” are imperative in ensuring the reproducibility and clarity of the guideline development process—a factor the previous recommendation development process lacked.

52.3.1 Transparency

A transparent guideline serves two main purposes:

The first is to ensure unambiguous and reproducible guidelines. Transparency ensures the guideline is clear and easy to follow. The treatment pathway should be well articulated.

A transparent guideline also fully discloses author information, conflicts of interest (COIs), and guideline funding.

Transparency allows physicians and patients to evaluate for themselves the reliability of and potential biases within a guideline. Ideally, this translucent nature of the guideline development process will deter biases from crossing into the development process, further cementing the trustworthy quality of the guideline.

52.3.2 Management of Conflicts of Interest

Any association between a developer and the guideline in question serves as a potential for conflicting interests. These associations may include academic interests, professional gains, personal gains, or financial advancements. Biases may cause the guideline to be developed unduly—Intentionally or not. Therefore, COIs must be limited to maintain guideline credibility.

To manage COIs, each individual participating in guideline development must disclose interests and medical and financial associations relating to the guideline. Ideally, disclosures work to minimize the biases that may seep into the development process assuring that the guidelines were not developed to suit certain interests while harming others.

52.3.3 Development Group Diversity

A trustworthy CPG depends greatly on its team of developers. A diverse team—One that includes a member from every discipline or party associated with its implementation or consumption—Can provide a well-rounded guideline with the interest of all parties protected [4]. This team includes primary care physicians, specialists, nurses, other providers, and any other party who may utilize the guideline. Patients, or other proxies, who may advocate for patients, must be also present. Patients and other proxies need no prior medical experience with the guideline topic as their role is to provide a voice for patients.

Such a diverse group ensures that patients’ needs are protected and concerns are respected.

52.3.4 Systematic Review

The systematic review (SR) process determines the inclusion and exclusion criteria for the literature search. During the SR process, articles are gathered, analyzed, and interpreted, and relevant data is summarized. The SR process begins with an all-inclusive search of the medical literature and ends with a preliminary draft of the guideline.

52.3.5 Evidence and Recommendation Strength

A ranking of the quality of evidence and the strength of research is conducted to apply appropriate weight to each guideline recommendation [20]. CPG developers focus on high-quality evidence to build their recommendation. Steering clear of overdependence on expert opinion is important as expert opinion may not be based upon well-rounded experience or complete information [4]. Basing guidelines on research with weak design or flawed methodology will result in biased or faulty guidelines. To ensure quality evidence, a quality assessment is performed for all research included in the guideline development.

Another factor the guideline team must consider while making recommendations is the fact that *a statistically significant finding may not be clinically relevant* [21]. To resolve this discrepancy, the America Academy of Orthopaedic Surgeons (AAOS) has applied the minimal clinically important improvement (MCII) method for determining clinical significance in research. This is similar to the minimally important difference (MID) or the smallest amount of change a patient may distinguish. Identifying clinical significance is important because a research finding that may be statistically significant to a researcher may not be relevant to patient treatment. Thus, certain research findings may not actually bear enough clinical weight to warrant a change in clinical treatment.

A statistically significant finding may not be clinically relevant [21].

For example, a pharmaceutical company has two drugs, Drug A and Drug B, which both decrease anxiety. Drug A has a success rate of 95%, whereas Drug B has a success rate of 89%. Drug A costs five times as much as Drug B and has much more severe side effects than Drug B. Statistically, Drug A has a significantly greater success rate; however, clinically, Drug B is far more appealing in the eyes of the clinician and the patient. It saves the patient money and potentially harmful side effects and yields nearly the same treatment outcome. In this case, the statistically significant finding isn't applicable in the clinical setting as the patient wouldn't be able to distinguish a difference between the successes of both drugs.

Medical literature is analyzed and ranked by its quality of study design—the highest-quality evidence corresponds to the lowest risk of bias.

The AAOS has developed a reliable “Clinical Practice Guideline Strength of Recommendation” rubric (Table 52.1) [3] that has been proven to generate strong CPGs. The

Table 52.1 AAOS strength of recommendation description table [23]

Strength	Overall strength of evidence	Description of evidence strength	Strength visual
Strong	Strong	Evidence from two or more “high”-strength studies with consistent findings for recommending for or against the intervention	
Moderate	Moderate	Evidence from two or more “moderate”-strength studies with consistent findings or evidence from a single “high”-quality study for recommending for or against the intervention	
Limited	Low-strength evidence or conflicting evidence	Evidence from one or more “low”-strength studies with consistent findings or evidence from a single moderate-strength study for recommending for or against the intervention or diagnostic test or the evidence is insufficient or conflicting and does not allow a recommendation for or against the intervention	
Consensus	No evidence	There is no supporting evidence. In the absence of reliable evidence, the work group is making a recommendation based on their clinical opinion. Consensus recommendations can only be created when not establishing a recommendation could have catastrophic consequences	

Table 52.2 AAOS recommendation language table

Guideline language	Strength of recommendation
Strong evidence supports that the practitioner should/should not do X, because...	Strong
Moderate evidence supports that the practitioner could/could not do X, because...	Moderate
Limited evidence supports that the practitioner might/might not do X, because...	Limited
In the absence of reliable evidence, it is the <i>opinion</i> of this work group that...	Consensus ^a

^aConsensus-based recommendations are made according to specific criteria. These criteria can be found in Appendix VII

strength of a recommended treatment pathway in a CPG is based off the quality of its supporting evidence (Table 52.2) [3]. Evidence quality is based on the following hierarchy of study design [3]:

- High quality: <2 study design flaws.
- Moderate quality: ≥2 and <4 study design flaws.
- Low quality: ≥4 and <6 study design flaws.
- Very low quality: ≥6 study design flaws.

Two or more high-quality studies yield a strong recommendation. One high-quality study or two or more moderate-quality studies yield a moderate-strength recommendation. One moderate-quality study and/or two or more low-quality studies yield a limited-strength recommendation. If there is conflicting evidence, the recommendation is ranked as limited [3, 22]. If there is no evidence to support the recommendation and the development team produces a recommendation, it is labelled “consensus” and is published in a separate companion consensus statement document to ensure separation between evidence-based and consensus-based recommendations. A “consensus” recommendation is equivalent to the historic expert group-based recommendation.

The terminologies “strong,” “moderate,” “limited,” and “consensus” are used to express the strength of recommendation within the guideline. After evidence analysis and recommendation ranking have occurred, the guideline is drafted.

In many cases, more than one recommendation may be presented. In instances such as these,

care maps or flow diagrams may be the best way to convey information as they are concise and easy to read and understand even when multiple variables are present. Multiple recommendations are included in the guideline to account for the variances that may arise. No two clinical cases are the same; as such the guideline provides a myriad of recommendations, so the physician may alter treatment as needed.

52.3.6 Articulation of Recommendations

Recommendations must be clearly written. They must be presented in the standardized format that includes a detailed treatment pathway as well as circumstances in which each recommendation should be used. Particular language is used to properly express the strength of the recommendation as well as the level of confidence the development team has in the recommendation. This information is vital as it allows the reader to evaluate how closely the guideline should be followed.

52.3.7 External Review

After a guideline is developed by the work group, but before it is released, an external review is conducted by an independent peer review group [4]. The external review serves as an independent, non-biased evaluation of the guideline. The group consists of medical professionals in related areas, persons from medical societies, and persons from the community. Just as the development team members are required to disclose COIs, so are the external review group members.

Reviewers are asked to review the evidence and comment on the wording of the recommendations. The peer review group is responsible for ensuring three main qualities of the guideline: validity, reliability, and feasibility.

1. *Valid* guidelines clearly state the scientific evidence supporting their recommendation. Justifications are present where group consensus and expert opinion were needed to support recommendation.

2. *Reliable* guidelines are *reproducible*. They are guidelines in which a peer reviewer comes to the same conclusion as the focus group.
3. *Feasible* guidelines are easily understood by both patients and physicians and allow for both routine use and case-by-case modifications when necessary.

The review team's written comments are collected into a single response form which is then reviewed and responded to by the chair of the guideline development team. Guideline development team members vote on all suggested revisions to recommendation language and are accepted with a majority vote [24]. The revision process is documented and reported in the guideline document until final guideline approval [24].

52.3.8 Updating

CPGs are subject to routine updates and amendments as new information presents itself or as time passes. Certain branches of medicine, such as the American Academy of Orthopaedic Surgeons and American Association of Otolaryngology-Head and Neck Surgery (AAO-HNS), update their guidelines at a minimum of 5 years after publication [21, 25]. Situations that warrant guideline updating may include but are not limited to [4, 26]:

- Changes/advancements in available treatment or intervention methods.
- New evidence that impacts current treatment.
- Changes in health-care availability, affordability, or access.

In addition to updates, CPGs may undergo amendments. There are three types of amendments:

1. **Reaffirmations:** This simply consists of a brief statement of the organization's agreement with the current guideline. This occurs when the guideline requires no significant alterations such as when time passes but treatment methods and research findings have not changed.
2. **Minor revisions:** A minor revision includes any alteration to the guideline that doesn't

change the overarching treatment plan but changes minor steps. This may be due to new research findings.

3. **Major revisions:** A major revision is any revision that significantly alters the treatment plan, course of action, or main conclusion of guideline.

Updates and amendments undergo an independent review and majority vote and are then published and distributed with an alert that the guideline has been revised.

Each guideline is accompanied by its "profile." A short statement that discloses the entire decision-making process includes the development team's values, the evidence quality and harm-benefit assessments, and the level of confidence the team has in the evidence. Limitations of the guideline are also expressed such as intentional vagueness the team may have included [4]. CPGs will not always be correct, as they must be revised with new information as research is published, nor will they be entirely bias-free. The CPG process outlined by the IOM aims to limit the amount of bias that seeps into the recommendation development process, increase consistency within patient care, and streamline the health-care delivery process.

Often a brief disclaimer is added to the beginning of the guideline abstract, such as this one from the American Academy of Otolaryngology-Head and Neck Surgery Foundation (AAO-HNSF) [4]:

This clinical practice guideline is not intended as a sole source of guidance in managing [topic specified here]. Rather, it is designed to assist clinicians by providing an evidence-based framework for decision-making strategies. The guideline is not intended to replace clinical judgment or establish a protocol for all individuals with this condition, and may not provide the only appropriate approach to diagnosing the managing the problem.

AAOS includes similar language in the introduction to the guidelines:

This guideline should not be construed as including all proper methods of care or excluding methods of care reasonably directed to obtaining the same results. The ultimate judgment regarding any specific procedure or treatment must be made considering all circumstances presented by the patient and the needs and resources particular to the locality or institution.

52.3.9 Implementation

Guidelines are only as effective as those who implement them. The National Guideline Clearinghouse (NGC) is responsible for the announcement, promotion, and distribution of CPGs [14]. Once the guideline is ready for implementation, it is crucial that physicians and all health-care providers use the guideline to deliver the highest quality of care to the patient.

52.3.10 Outcomes Assessment

Outcomes assessments are important measures to determine whether or not treatment has been successful [14]. A CPG, like any treatment, undergoes an “outcome assessment.” However, the organizations typically involved in outcome assessments, such as the National Quality Measures Clearinghouse (NQMC) and others, are not involved in the outcome assessment of CPGs. This is because the evaluation of outcomes is built into the CPG development process and expressed by the ranking of the recommendations. As such, the IOM Committee on Quality Health Care in America resolves that there is no need for the rating of quality measure of CPGs as it would be redundant. Moreover, an additional rating could create conflicts of interest as some CPG developers also develop related outcome assessment rubrics [14].

Clinical Vignette

Management of Anterior Cruciate Ligament Injuries:

As of 2017, the AAOS has completed 18 CPGs. In 2015, the AAOS published a guideline for the “Management of Anterior Cruciate Ligament Injuries” providing physicians with a detailed, outlined plan to aid in the prompt and accurate treatment of ACL injuries [27].

The topic was chosen for guideline development as some controversy existed over best treatment and management

options, and this condition impacts a significant number of patients in the USA. The extensive literature had not yet been concisely accumulated to distinguish best treatment options for appropriate cases. A systematic review of the literature was conducted in adherence with the aforementioned guideline criteria. In the ranking of their guideline recommendations, the guidelines were assigned a star grade to easily distinguish strength of recommendation (4 stars for a strong recommendation, 3 stars for a moderate-strength recommendation, etc. in accordance with Table 52.1). The stars aligned the IOM’s “strong,” “moderate,” “limited,” and “consensus” rubric language terms.

Of the 20 recommendations put forth by this ACL injury management CPG, 5 had strong supporting evidence (4 stars), 6 had moderate supporting evidence (3 stars), 7 had limited supporting evidence (2 stars), and only 2 were consensus-based. This CPG shows not only great advancement in the strength of the orthopedic research being conducted but also great improvement in the guideline themselves. Only 6 years prior, in 2009, AAOS published its first CPG on diaphyseal femur fractures in pediatrics. This guideline, although much stronger than its consensus-based predecessor, only had 1 recommendation out of its 14 that would have received a 4-star ranking and only 2 that would have received 3 stars. The 2015 CPG on ACL injury management shows great advances from both an orthopedic research perspective and a clinical practice guideline and care management perspective.

The following are a few examples of the strongest recommendations from the ACL injury management CPG [27]:

1. “Strong evidence supports that the practitioner should obtain a relevant history and perform a musculoskeletal exami-

nation of the lower extremities, because these are effective diagnostic tools for ACL injury” (4-star/strong evidence recommendation).

2. “Strong evidence supports that the MRI can provide confirmation of ACL injury and assist in identifying concomitant knee pathology such as other ligament, meniscal, or articular cartilage injury” (4-star/strong evidence recommendation).
3. “When ACL reconstruction is indicated, moderate evidence supports reconstruction within five months of injury to protect the articular cartilage and menisci” (3-star/moderate evidence recommendation).

As the AAOS has gained more experience with the guideline process, and the overall quality of the orthopedic literature has improved, more recent guidelines include questions that follow patients through the path of care. The language for recommendations reflects the quality of evidence in research publications. The recent guidelines for management of elderly hip fractures and ACL injury are examples of high-level guideline recommendations.

The language for recommendations reflects the level of evidence in research publications. The recent guidelines for management of elderly hip fractures and ACL injury are examples of high-level guideline recommendations.

52.4 Turning CPGs into a Care Map

CPGs provide clinical guidance for a wide variety of topics and help manage specific conditions. CPGs may be used to support care maps,

which can help local medical groups and health system and outline care pathways. Simple, easy-to-follow flow diagrams help providers quickly determine the best route of care for each specific clinical case.

The AAOS 2014 CPG for developmental dysplasia of the hip (DDH) was used to develop DDH Care Map for the St. Luke’s Health System in Idaho [28]. The CPG summarized recent research and clinical treatment options for the evaluation of DDH, creating evidence-based treatment options for different clinical presentations of DDH. This guideline was turned into an easy-to-follow care map, providing practitioners with step-by-step instructions on how to best treat each specific case. The map includes treatment methods varying with patient age and degree of hip dysplasia. It also includes modifications to treatment for those clinics that may not have access to the ideal imaging machines. Moreover, it is easily accessed via smartphone, tablet, iPad, or other portable screens providing ease of access for providers and families.

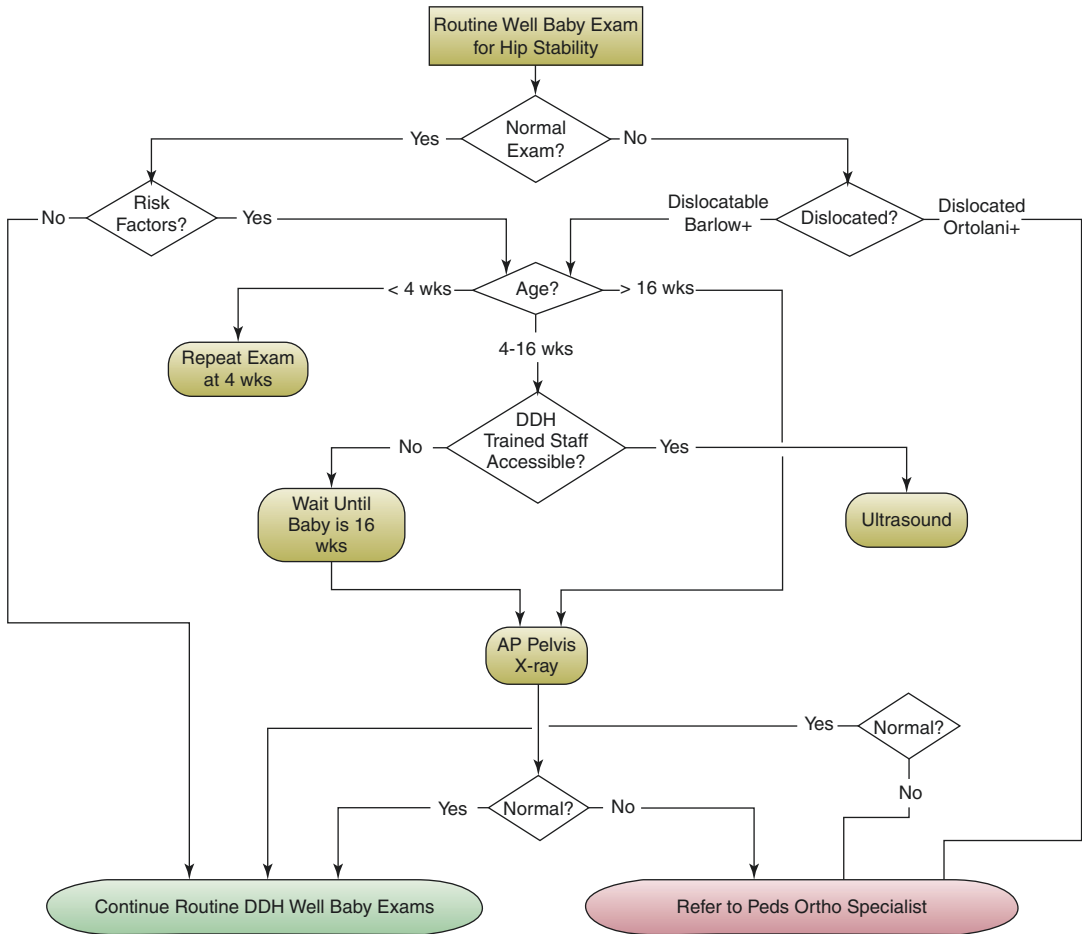
The DDH Care Map (Fig. 52.1) continues to be used in clinical practice in the St. Luke’s Health System, and this care map is continually reviewed and updated. Feedback from clinicians, as well recent publications, will lead to changes in the care map.

AAOS CPG program led to the development of other care maps for other health systems and practices, including those for carpal tunnel syndrome and for management of elderly hip fractures.

52.5 Limitations of CPGs

Many areas of medicine, like orthopedics, have massive medical literature databases. Sorting through the extensive databases for appropriate articles and ranking levels of evidence to develop CPGs requires considerable effort and expertise [3]. CPGs are beneficial for many reasons; however, some disadvantages include:

1. The process is time-consuming and expensive.
2. Patient feedback is ideal, but often not available from in the literature published.



Diagnosis Screening and Referral Pathway

Ultrasound is the preferred imaging study until 6 months of age. Radiographs are indicated thereafter. If ultrasound is unavailable, radiographs can be used as early as 3 months.

Fig. 52.1 Care map for DDH [29]

3. Guidelines are subject to misinterpretation.
4. Guidelines must be continuously updated, to reflect changes in the published literature.
5. Adequate literature is required to develop CPGs, so areas lacking in literature won't qualify for CPG development.
6. Guidelines are only as effective as those who implement and abide by them.
7. For less common conditions, adequate research literature may not support the development of a CPG.

For those guideline patient care questions that lack relevant research or have an insufficient

database, members of the clinical practice guideline work groups are allowed to create a companion consensus statement [14]. These are statements based on expert opinion and are published separately from the CPGs to ensure separation of expert-opinion-based recommendations and evidence-based recommendations.

52.6 Future Studies

Although the guidelines have limitations, the guidelines are beneficial for many reasons. The guideline process identifies medical areas which

lack higher levels of research and highlight the direction for future research. A lack of evidence to develop strong evidenced-based CPGs is common in many medical specialties. A review article evaluating the strength of over 2700 recommendations put forward by the American College of Cardiology and the American Heart Association found that only 11% of those 2700 recommendations were based on Grade A, or “strong,” evidence [3]. CPGs may not always be feasible to produce as certain topics lack sufficient data, but they do provide a service—highlighting important gaps in research and important clinical questions that must be answered in to provide optimal patient care.

Take-Home Message

- The IOM and others have called for a revision of the development of the highest standards of care in the American health system.
- Evidence-based clinical practice guidelines have been designed to replace the consensus-based guideline to increase health-care efficiency and patient care success.
- Guidelines have limitations, but they can have a positive impact on patient care.
- They are designed to streamline patient care—potentially aiding in the treatment of the anticipated increase of chronic conditions.
- As CPGs serve as a summary of scientific evidence available, those areas which lack adequate clinical research may become research priority.
- Through the extensive and deliberate analysis of high-quality medical literature, CPGs provide evidence-supported health-care plans for physicians and patients alike.
- Ultimately, these guidelines may reduce practice variation, improve quality of care, decrease inefficiencies, and withstand the scrutiny the previous guideline process could not.

52.7 Useful Websites

<https://www.aaos.org/guidelines/?ssopc=1>
<http://www.orthoguidelines.org/topic?id=1018>

References

1. Graham R, Mancher M, Miller Wolman D, Greenfield S, Steinberg E, eds. *Clinical Practice Guidelines We Can Trust*. Washington, DC; 2011. <https://doi.org/10.17226/13058>.
2. Shea KG, Sink EL, Jacobs JC Jr. Clinical practice guidelines and guideline development. *J Pediatr Orthop*. 2012;32(Suppl 2):S95–100.
3. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC; 2001. <https://doi.org/10.17226/10027>.
4. Rosenfeld RM, Shiffman RN, Robertson P, Department of Otolaryngology State University of New York. *Clinical practice guideline development manual, third edition: a quality-driven approach for translating evidence into action*. *Otolaryngol Head Neck Surg*. 2013;148:S1–55.
5. Foundation TRWJ. *Chronic care in America: the 21st century challenge*. 1996. <http://www.rwjf.org/library/chrcare/>. Accessed 19 Sept 2000.
6. Chassin MR, Galvin RW. The urgent need to improve health care quality. *Institute of Medicine National Roundtable on Health Care Quality*. *JAMA*. 1998;280:1000–5.
7. Cooke CR, Gould MK. Advancing clinical practice and policy through guidelines: the role of the American Thoracic Society. *Am J Respir Crit Care Med*. 2013;187:910–4.
8. Hoffman C, Rice D, Sung HY. Persons with chronic conditions. Their prevalence and costs. *JAMA*. 1996;276:1473–9.
9. Gerteis J. *Multiple chronic conditions Chartbook: 2010 medical expenditure panel survey data*. 2014. www.ahrq.gov/sites/default/files/wysiwyg/professionals/prevention-chronic-care/desicion/mcc/mccchartbook.pdf.
10. Hempstead K. The real killer is still out there: update on health care spending. 2017. <http://www.rwjf.org/en/library/research/2017/08/the-real-killer-is-still-out-there-update-on-health-care-spending.html>.
11. Roehrig C. The health SPending slowdown for 2008–2013: implications for sustainability. 2017. http://altarum.org/sites/default/files/uploaded-related-files/11_Roehrig%20Symposium%20July%2018%202017%20FINAL.pdf.
12. Centers for Disease Control. *Arthritis—National Statistics*. 2017. Located at: CDC.
13. Association PlaAH. *Quality and patient safety*. 2006. <http://aha.org/advocacy-issues/quality/background/shtml>.
14. AAOS. *Clinical Practice Guidelines*. *Am Acad Orthop Surg*. 2017.
15. Als-Nielsen B, Chen W, Gluud C, Kjaergard LL. Association of funding and conclusions in randomized drug trials: a reflection of treatment effect or adverse events? *JAMA*. 2003;290:921–8.
16. Bekelman JE, Li Y, Gross CP. Scope and impact of financial conflicts of interest in biomedical research: a systematic review. *JAMA*. 2003;289:454–65.

17. Riaz H, Raza S, Khan MS, Riaz IB, Krasuski RA. Impact of funding source on clinical trial results including cardiovascular outcome trials. *Am J Cardiol*. 2015;116:1944–7.
18. Gagnon A. Corporate influence over clinical research: considering the alternatives. *Prescrire Int*. 2012;21:191–4.
19. Turner EH, Matthews AM, Linardatos E, Tell RA, Rosenthal R. Selective publication of antidepressant trials and its influence on apparent efficacy. *N Engl J Med*. 2008;358:252–60.
20. Shea KGMC, Quinn R, Beckmann JT. Evidence based quality and outcomes assessment in Pediatric Orthopedics. Rosemont, IL: AAOS; 2016.
21. Jevsevar D, Shea K, Cummins D, Murray J, Sanders J. Recent changes in the AAOS evidence-based clinical practice guidelines process. *J Bone Joint Surg Am*. 2014;96:1740–1.
22. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336:924–6.
23. American Academy of Orthopaedic Surgeons. AAOS Introductory Packet for Clinical Practice Guidelines (CPG)/Systematic Review (SR) Work Group Members. 2015:31.
24. AAOS. Guideline peer review and public responses. <https://www.aaos.org/guidelinepeerreview/?ssopc=1>. Accessed 20 Nov 2017.
25. Waters E. Evidence for public health decision-making: towards reliable synthesis. *Bull World Health Organ*. 2009;87:164.
26. Shekelle P, Eccles MP, Grimshaw JM, Woolf SH. When should clinical guidelines be updated? *BMJ*. 2001;323:155–7.
27. American Academy of Orthopaedic Surgeons. Management of anterior cruciate ligament injuries. Located at: Clinical Practice Guidelines. 2015.
28. Kevin G Shea CTP. Turning a CPG into a care map. 2016. <https://www.aaos.org/AAOSNow/2016/Oct/Research/research03/>. Accessed 14 Nov 2017.
29. AAOS. DDH Care Map-Diagnoses and Referral Pathway. 2016. <http://hipdysplasia.org/diagnosis-and-referral-pathway/>.

How to Navigate a Scientific Meeting and Make It Worthwhile? A Guide for Young Orthopedic Surgeons

Darren de SA, Jayson Lian, Conor I. Murphy,
Ravi Vaswani, and Volker Musahl

53.1 Introduction

“When minds meet, they don’t just exchange facts; they transform them, reshape them, draw different implications from them, and engage in new trains of thought.

Conversation doesn’t just reshuffle the cards, it creates new cards”

Theodore Zeldin (b.1933) Historian & Author

Scientific meetings present immense potential to enhance one’s overall well-being in both personal and professional realms. Often a welcomed break and change in tempo from the hectic clinical schedule, these meetings present opportunities to not only reconnect with peers and/or establish new contacts but also to focus on evidence-based practice. Given that the quantity of scientific literature continues to expand at an exponential rate, and considering recent evidence suggesting the poor short-term publication rate of posters and podiums presented at various meetings [7–9], attendance and more so active participation at these meetings enable one to remain “ahead of the curve” in their chosen field. Though at the very least, these are

opportunities to explore new geographic regions and cultures, they provide ample opportunity to refresh, reignite our passions, avoid burnout, and ultimately, deliver high-quality care. To fully take advantage of a *scientific meeting*, thorough preparation is warranted, with success akin to a rigorous “preoperative plan” that has been visualized, rehearsed, and ready to adapt to sudden changes. This chapter, modeled after the “preoperative plan,” will outline some key elements of a *scientific meeting* and present a helpful guide for pre-, during, and post-meeting actions to maximize the personal and professional impact.

53.2 Clinical History

Well before one sets foot in a conference venue, significant preparation, often at a minimum of 2–3 months in advance, is required. Although not always obvious, the key first step is to identify a meeting of professional relevance among the numerous listed to maximize time-value efficiency. It is wise to often consult colleagues and mentors for recommendations, to examine previous conference programs, and to avoid common pitfalls of attending a meeting simply to keep the status quo (i.e., “everyone is attending”), or because of its sheer size (i.e., “bigger is not always better”). Though this holds true of any meeting, it is of importance especially if a first-time attendee to prioritize meetings based on an

D. de SA · C. I. Murphy · R. Vaswani
V. Musahl (✉)

Department of Orthopaedic Surgery, University of
Pittsburgh Medical Center, Pittsburgh, PA, USA
e-mail: musahlv@upmc.edu

J. Lian
Albert Einstein College of Medicine,
Bronx, NY, USA

Fact Box 53.1

SMART goals are specific, measurable, achievable, results-focused, and time-bound. They are a recognized method of improving efficiency and thoroughness of task completion [3]. If used when planning one's approach to a *scientific meeting*, they can help the attendee make the most of the experience.

ability to achieve personal goals (i.e., education, *networking*, visiting/trialing new *vendor* products, marketing, etc.). Know the conference scientific program and its intended audience, and from this, establish SMART goals (see Fact Box 53.1) for both the immediate conference and short-term period thereafter—that is, objectives that are specific, measurable, achievable, results-focused, and time-bound [3]. Remember, depending on the size of the meeting, its individual components can range quite broadly and include, but are not limited to, didactic lectures, small-group discussions/workshops, multiple *breakout* sessions of smaller focus, various scientific and *vendor* exhibits, and numerous social activities.

It is not necessary, or often not feasible, to do it all.

Upon selecting a meeting of interest, know the important dates (i.e., “early bird” registration, calls for abstracts, award application deadlines, accommodation booking dates, deadlines for registration of pre-courses or supplemental activities, etc.). Whenever possible, consider submitting content, be it abstracts, *instructional course lectures (ICLs)*, video demonstrations, or the like, to optimize chances of being an active participant, engaged in the meeting. This will best ensure that information is not only obtained but further retained with higher likelihood of being practically applied. At the very least, add your name to the conference mailing list, to keep abreast of updates/latest information.

The importance of being fiscally responsible has not escaped notice. To this end, search for institutional and conference-specific funding sources to support travel and participation, and be

cognizant of the cost-savings from “early” registration and accommodation booking. Though often influenced at an individual institutional level, *junior learners* are often provided with limited access to funding, and so early arrangements are particularly beneficial for this group. Regardless, it is highly recommended to consult with travel agencies in collaboration with a meeting, to ensure cost-effective and effortless arrangements are made. This is particularly true if bringing family, as it is recommended to not only consult the conference program but also the visitor center for the conference location to develop an itinerary of attractions and restaurants to keep your companions entertained.

53.3 Physical Exam

Know the meeting, well in advance. This is an opportunity to see if the individual components of the meeting align with one's understanding of its intended aims and audience. Review, in-depth, the scientific program; download meeting-specific smartphone applications; and note dates of interest (i.e., Welcome Reception, Opening Ceremony, Guest/Keynote Speaker, meeting highlights). Knowledge of the different events is important, facilitating one to pack comfortable, yet appropriate, professional clothing for any setting: delivering a *podium presentation*, serving as part of a *panel discussion*, attending a “black tie” fundraiser or “cocktail reception,” etc. Moreover, as meetings are shifting toward encompassing more active audience participation—often mediated through smartphone applications—obtaining and testing the smartphone application ahead of time will enable one to be savvy with navigating throughout the meeting and participating in a timely fashion. Though it may appear daunting, consider reading the accepted abstracts for *podiums* and posters of sections of personal interest, note the authors, review the faculty profiles, and compose, ahead of time and where applicable, one or two questions of significance to ask should the opportunity arise. Preparation is key, and this advanced preparation will undoubtedly enhance understanding of conference content and facili-

tate *networking* as well. To this end, armed with this intimate knowledge of conference content, it would be wise to contact and establish meetings with people of interest at this juncture, well in advance of the meeting date, as it can be nearly impossible to exchange itineraries at the venue itself. In-depth review of the conference outline and content also will facilitate advanced preparation for technical workshops and maximize educational yield. It will enable one to contribute to as well as gain from interactions with others in these more focused sessions.

53.4 Imaging

Though especially important for the large society annual meetings, reviewing the conference and accommodation venue layouts is paramount. Obtain the “big picture” of where large group (didactic) sessions are located versus such other elements as *symposia*, *podium presentations*, poster/e-poster presentations, surgical demonstrations, *instructional course lectures*, *panel discussions/debates*, *technical/vendor* exhibits, hands-on workshops, career development/practice management booths, etc. so that such knowledge of both the content and layout will enable thoughtful planning of an individualized schedule. Know when scheduled breaks are and take them. Do not plan on filling every moment of the day with conference-related activity, and make a conscious effort to sample all the different types of settings for delivering content, all while balancing scheduled personal time to focus on fatigue management [22] and sleep hygiene [4].

53.5 The Approach

Arrive early, well rested, and free from outside distractions. A “pearl,” if possible, would be to not bring unfinished work (i.e., dictations, manuscripts to review, etc.) to the meeting. While in attendance, one should be wholeheartedly focused on being mentally and physically present. If necessary, perform a pre-meeting “time-out” to again review one’s intended and individualized goals,

the timeline, and plan for execution to maintain focus. However, understand that the pre-meeting plan is not “written in stone” [5], and those who truly benefit from meeting attendance often spend time to reflect at the end of each conference day on the value obtained from the pre-plan—maintaining flexibility to modify the remainder of the schedule to maximize experiences and follow-up on new connections made. Again, thorough knowledge of all opportunities a given meeting presents ahead of time eases this transition. Typically, during the check-in process, participants receive materials including the full program, an identification badge, and a tote bag. It is important at this juncture to identify key locations including the resource center, exhibit hall, lecture hall, and food and restroom locations and obtain Wi-Fi access at the conference location.

53.5.1 Education

While visiting *symposia*, *podium presentations*, and poster exhibitions, among other conference events, it is critical to be engaged. However, being engaged oftentimes does not lead to retention of learned information at the conference [17]. In fact, 3 and 90 days following a conference, the mean retention rate was reported of only 14.9% and 11.3%, respectively [17]. One solution for this may be purposeful self-selection and attendance of poster presentations or *podium presentations*. Secondly, having access to conference

Fact Box 53.2

Flipped sessions involve exposing learners to content before the classroom session, so that learners are more prepared and the classroom sessions can focus on interactive activities and questions the learners may have [13]. This model works well for *scientific meeting* sessions as learners have varying levels of experience, so the learners can tailor the sessions to fit their needs. The moderators of the sessions can then act as facilitators rather than lecturers [13].

materials in advance and preparing for meaningful dialogue can promote and enhance a “flipped session” format, which is a contemporary model of learning in which lecturer and viewers engage one another with question and answer—shown to aid learning and retention [13, 21]. As more conferences adapt this format [13, 21], preparation for conferences can greatly augment the conference experience. That being said, approach the conference as a learner, and set aside personal agenda and/or ego.

Conferences are significant in that current and up-to-date concepts are discussed among experts in the field. During meetings, listening to “buzz-words” or “themes” may help flavor future research directions as they provide a “snapshot” of the field. Likewise, learning about ongoing, impactful research projects generates new research ideas and innovation. Often, new ideas are invented during conversation among orthopedic surgeons at these conferences.

As the concept of a *scientific meeting* evolves, so too does technology, and as such, the internet has drastically changed the way in which conference participants interact. An increased number of conferences are utilizing social media, such as Twitter® or smartphone meeting/conference applications, to engage conference participants as well [14]. For example, there are times when traditional poster presentations fail to elicit engaging questions and enthusiastic viewers. Traditional oral presentations can largely be lecture style and less interactive. Some conferences have therefore utilized Twitter and other social media outlets to encourage conversation among conference participants and interested outsiders [14]. In a study among the urological community, for example, conference attendees using Twitter® found it beneficial for *networking* (97%), spreading information (96%), research (75%), advocacy (74%), and career development (62%) [1]. As the social media presence of orthopedic surgeons grows [18], the importance of utilizing Internet platforms to bolster conversation and dialogue cannot be ignored.

The Internet, however, can be a “double-edged sword,” and though it poses tremendous advantages

to a conference participant, it is equally critical that it not be used to answer work-related emails or provide means to other distractions. This only detracts from the purpose and opportunity of the conference. Furthermore, compartmentalization of work-related materials to the end of the day, and focusing on it with full attention is a strategy shown to provide improved efficiency and results [11].

53.5.2 If Presenting

Preparation is of the utmost importance for executing a successful presentation at a *scientific meeting*. Initially it is important to research effective presenting strategies and techniques in anticipation of the presentation. Constructing slides that are readable and concise in a format that is easily consumed by the audience is difficult and must be consistently at the forefront of the presenter’s mind during composition. Formatting and color schemes need to be consistent with other presentations within the department or laboratory where the work was performed. Furthermore, formatting should adhere to any listed guidelines for the meeting. Standard content includes a brief background, hypothesis and aims, methodology, results, conclusions, and discussion. Each slide should be purposeful, with a “take-home message” that is easy to convey. Use of images can be particularly helpful. Lastly, the slides and material must be tailored to the targeted audience. Attendees may be comprised of scientists, clinicians, *vendors*, mathematicians, statisticians, engineers, residents and fellows, or any combination thereof. The emphasis and message of the presentation may change depending on the target audience.

After the slides have been prepared and edited by the presenter and advisor, live practice to hone timing and delivery and anticipate potential audience questions is necessary. There should be no difficulty complying with time limits established under the presentation guidelines. Failure to do so demonstrates lack of preparation and will distract from the presentation and “take-home messages.”

Scripting the presentation so as to avoid unnecessary pauses or lapses is helpful. Lab meetings or educational conferences are advantageous venues to practice in front an audience that will be similar to the *scientific meeting*. Monitor the timing, elicit feedback, and make final edits as needed to clarify the message of the presentation. Remember, you represent not only yourself but your institution—and this must not be overlooked.

At the *scientific meeting*, the first manner of order is to locate the “Speaker Ready” room to upload presentation materials. Most meetings will specify a location with technical assistants, ready to upload the appropriate materials. Follow the listed instructions to avoid any difficulties. Maintain a backup of the presentation on a portable memory device and/or in an e-mail inbox, should the upload encounter problems.

Finally, visit the venue in advance of the presentation in order to become familiar with the room and the technical tools available. Ensure that any pointers, clickers, or other presentation tools are in working order. It can be helpful to attend another presentation in the same venue earlier, to observe other technical shortcomings or mistakes so as to avoid repeating them. It can also help a great deal with any degree of public speaking anxiety.

53.5.2.1 How to Make the Most of Roundtable Discussions?

A *roundtable* provides a small group of participants (typically 10–12) the opportunity to, over a short 60–90 minute time period, partake in a group discussion across specific items of a topic of interest. Given its academic nature, a *roundtable* encompasses a rapid exchange of high-level information among peers, with its success resting on input from all members in attendance. Often, it is an opportunity to clarify approaches to current problems and provide an avenue to present and weigh all viewpoints. Therefore, it is paramount that you first attend a *roundtable* that not only is of particular interest to you but to which you have a sufficient knowledgebase and that is being attended by peers, ideally those you are familiar with, so that contributions can, at the very least,

be delivered with vigor and glean equal insights. Although often led by a moderator to keep to a particular task at hand, it is important to remember the group nature of a *roundtable*, and as such, one must ensure that when participating, all members present are being addressed and afforded opportunities to engage as well. Questions should be posed with appropriate tone, so as to convey a sense of open-mindedness and to be mindful of the process. To that end, all action items of the *roundtable* may not be addressed during the session, and emphasis should remain on thoroughly addressing a particular item, as opposed to rushing transitions through topics or pressing for a consensus resolution.

53.5.2.2 How to Make the Most of Panel Discussions?

On the contrary, a *panel discussion* involves members of the audience listening to the perspectives of experts who have been pre-selected to the panel. Each individual panel’s focus may be different, and as an audience member, it is important to realize that members of the panel have different roles, suited to the particular meeting. Not all panel members may agree with each other. Panel members may be pre-selected to present a viewpoint in keeping with their individual practice, to present a viewpoint contrary to their practice, to debate other panel members, etc. Thus, to maximize the yield from attending *panel discussions*, one should keep this in mind and anticipate a passive learning experience for the most part. It is helpful to not only prepare for the topic and anticipated viewpoints ahead of time but to also research background information on each panel member, to familiarize yourself with their prerogative in advance. This will further enable one to identify some key areas of controversy and possible areas that will be addressed. Though opportunities do exist to pose questions to the panel, it is advised to use this platform wisely, be mindful of the others in attendance, be respectful of everyone’s time, and not use this opportunity as an avenue to express your own perspectives, to debate the panelists, or to engage in lengthy one-on-one discussions.

53.5.2.3 How to Make the Most of Breakout Sessions/ Hands-on Workshops?

Breakout sessions, particularly those with a hands-on focus, provide a unique opportunity to integrate learning, challenge oneself, and learn from expert facilitators leading the workshop. Preparation is the key to maximizing return from these opportunities. Review the anatomy, practice with the equipment, review the procedural sequence, etc. ahead of time. One must possess strong self-reflective abilities and select workshops that are mindful of one's ability and level of training. Unless specified as introductory, or focused on a novel technique, one should thoroughly prepare for this session by preemptively obtaining the necessary experiences that can subsequently be built upon during the workshop. Hands-on workshops provide a great opportunity to refine foundational skills, learn new approaches to similar problems, exchange tips/tricks, and, in a safe environment, experiment with new techniques outside of their comfort zone. Practice makes permanent [2]; and so, while these are often artificial environments, it is important to visualize the clinical setting in which the skill is being applied, to simulate the real-world environment as much as possible. Poor preparation will prevent any meaningful skill acquisition and will be pointless for both yourself and the instructor(s).

53.5.2.4 How to Make the Most of Online Tools

The increasing utilization of technology within orthopedic surgery has also expanded into the realm of *scientific meetings*. Online tools continue to become more available to assist attendees navigate the meeting. First, the *scientific meeting* website operates as the primary tool for registration, description, updates, and information. Programs, relevant academic materials, speakers, maps, exhibitions, lectures, and other special events will all be listed here. It is important to visit the website well in advance of the meeting to register and pay any fees. This will accelerate the on-site registration and possibly avoid any late increases in price of registration as the meeting gets closer. The website will also

contain information for specific populations of registrants that may be available for scholarships or price reduction to attend the meeting. There will possibly be price reductions at local hotels, restaurants, or other local establishments for registrants as well.

Some meetings may have interactive agendas listing the most up-to-date schedules of events. In order to maximize time and efficiency at the meeting, it is important to study the upcoming daily agenda and plan out the day ahead of time in order to avoid missing lectures or exhibitions of personal interest and relevance. At larger meetings with several simultaneous events, the website and online agenda can be the best tool to help plan out a personal daily itinerary. It is often easy to sort events by topic or speaker in order to isolate specific interests.

Many *scientific meetings* publish a dedicated app for mobile devices. Depending on the meeting, this app may function as the primary tool for event feedback, enrollment in daily activities, voting in polls, or announcing last-minute changes in venue or speaker. Associated video lectures, learning modules, abstracts, posters, and referenced manuscripts can be available through mobile apps as well. Briefly reviewing these materials prior to attending the event greatly supplements presentations and augments retention. Prior to the meeting, check the website for announcements regarding available apps, and download it to your mobile device.

53.5.3 Networking

While *networking* strategies have been critically analyzed and developed in other specialties, it seems there is a paucity of literature in the medical community when approaching this topic. The reason for this is multifactorial; however, it is important to recognize that *networking* is an important skill all physicians should acquire. Like all skills, *networking* has both instinctive and learned components and can be improved over time, but it also can be prepared and trained. As medicine is largely founded on collaboration and teamwork, this should be paramount in

physician development and training. In this respect, conferences are key opportunities to meet and connect with other physicians with shared interests.

Nowadays, business cards are less important, and there are countless methods to network online, such as through LinkedIn®, ResearchGate®, Facebook®, Twitter®, and Instagram®. For example, among nearly 1000 Pediatric Orthopaedic Society of North America members, 95% of members had a professional webpage, 36.8% a LinkedIn® page, 33% at least one YouTube® video, 25.8% a ResearchGate® page, 14.8% a professional Facebook® page, and 2.2% a professional Twitter® page [10]. Among members, private-practice physicians had double the utilization of social media [20]. Growing an Internet presence prior to the conference may additionally improve *networking*.

As alluded to earlier, further preparation for *networking* can include identifying presenters that one may be interested in meeting, or professional societies one may be interested in joining. Performing the necessary research on their areas of interest, and finding a niche for potential collaboration, can increase the likelihood for successful *networking*. Preparing a three-sentence “Tell me about yourself” pitch should also be practiced, rehearsed, and delivered confidently. Recent trends suggest the content of these “elevator pitches” has shifted toward *why* a business exists and that the average adult attention span is currently just 8 s—perhaps due to the advent of smartphones [16]. Seizing the brief windows of opportunity that arises during a conference, with esteemed professionals in the field, can dramatically affect career path. As such, the importance of attending the social events, alumni and society meetings, and approaching people has not gone unnoticed. Besides addressing one’s nutritional needs, noting the breakfast and lunch schedules can be invaluable *networking* opportunities, and it is key that one sits at an active table, engages in discussion with others, and removes any distractions (i.e., concomitant laptop or smartphone use). Try to balance efforts on meeting new people with time spent with members from one’s own institution or previous colleagues at other

institutions as well, as the latter provides an opportunity to reconnect, reinforce relationships, and share experiences in a less-judgmental and/or anxiety-provoking environment.

For *junior learners* going to conferences, *networking* can come in a variety of forms: among peers from other institutions, with faculty, and with potential mentors. Mentors are invaluable resources for learners and can provide intangible benefits such as letters of recommendation, job positions, and career-long advice and guidance. *Junior learners* can use the meeting as a chance to spend time outside of the work environment with their mentors and build a more personal relationship. Furthermore, by following a senior faculty member, learners can also have the opportunity to meet and network with other senior members of the community. For mid-career and senior leaders, it is mutually beneficial to develop their mentorship capabilities. While it can be daunting for learners to network, it is simultaneously imperative that *junior learners* seize opportunity to learn from more experienced physicians and that they get sufficient face time with leaders in the field. This is especially important given the “Internet age,” as the younger generation is more likely to engage online rather than face to face [15]. *Junior learners* should also take notice of how more senior conference attendees act in order to learn and seek advice regarding the “unspoken rules of proper conference etiquette” [19]. Further, similar to resident and attending physicians, a *junior learner* can build his/her reputation in the field by their behavior at a conference. Lastly, finding a genuine connection to *junior learner* peers can go a long way, since these fellow learners will most likely continue to become peers and leaders in the field in the future.

53.5.4 Equipment/Vendors/Suppliers

Sponsors, *vendors*, equipment suppliers, and many medical companies occupy a large and central role in *scientific meetings*. Many attendees avoid these areas and interactions with representatives to avoid the stigmatized label of being an

“industry-sponsored physician.” The relationship between industry and physicians has been scrutinized and more regulated recently due to past unscrupulous practices by some physicians and industry leaders [6]. However, attendees should not be apprehensive about going to these areas of the *scientific meeting* as industry representatives play vital roles in the hospital to provide patient care and assist physicians. Establishing relationships and developing them over time have benefits for both sharpening surgical technique and providing enhanced patient care. Mastering the plethora of tools in the surgical armamentarium and recognizing the capabilities and deficiencies of each are an arduous process of experience. *Vendors* and representatives from equipment suppliers are a wealth of technical product knowledge and can be used to help expedite this learning curve. They can share “tips and tricks” from other surgeons and provide discounted access to educational opportunities such as cadaver labs, sawbones sessions, and research opportunities to compare products. Lastly, these connections may increase surgeon exposure to other techniques and equipment serving similar existing purposes to those at home institutions but at a decreased cost. As the landscape of medicine transitions to a value-based healthcare approach with an evolving financial payout methodology [12], reduced equipment costs with similar function and outcome can play a pivotal role.

53.6 Closure

At the conclusion of the *scientific meeting*, if done right, one should be refreshed, be reinvigorated in their chosen field, and possess a sense of accomplishment. The initial moments represent a vital period for both reflection and following up on connections established during the meeting. Though not necessary, any notes made should be revisited to reinforce key ideas, return to further reading materials/resources, or identify the springboard for the next set of personal activities. Be critical of your experience, evaluating what content and session format was most enjoyable/

productive and what formats resulted in disengagement. This is important to help form future conference planning, though a fair degree of tolerance should be exercised as poor experiences in one format may be one-off experiences due to a multitude of factors. Often meetings have post-conference access to the program, video presentations, copies of presentations, etc., and one should obtain copies of these for future reference materials.

53.7 Follow-Up

It is important to establish a platform to share the knowledge gained from the meeting with others, as this allows the reach of the particular conference to grow exponentially and further integrates the key concepts. Preferably, these summary meetings are established pre-departure. Lab members can collaborate by assigning certain topics beforehand to each member who is attending. Then, in the summary meeting, each member will summarize what he or she has learned for the rest of the group. This way, all members can expand their knowledge from the meeting even if they did not attend.

The importance of following up with key connections made during the meeting cannot be underestimated. Do not hesitate to implement social media, e-mail, etc. to send a short message to new connections—not only does this reinforce to them the value of their personal connection, but this may further lead to future collaborations and *networking* opportunities. Finally, if not already done, this remains a last opportunity to not only join the society hosting the meeting, if of interest, but to also provide meeting feedback and ensure continuing medical education (CME) certificates are received/statuses updated.

53.8 Conclusion

In conclusion, *scientific meetings* are excellent avenues for pursuing continuing medical education. To take advantage of these meetings, a

Fact Box 53.3

The online resources often include supplemental information and meeting schedules. These include MyAcademy App, aaos.org/annualmeeting, ORS app, isakos.com/meetings, My POSNA App, and ota.org/education/meetings-and-courses.

comprehensive researched strategy prior to arrival enhances the experience and allows the attendee to make the most of the time. The extra-curricular exhibits such as the *vendor* displays, technical demonstrations, and social engagements supplement the experience by providing opportunities for *networking* and establishing professional relationships. Multiple online tools such as websites and apps exist to streamline the dissemination of meeting information in a portable and easy-to-use fashion. Post-meeting analysis sessions help promulgate the energy and information shared at the *scientific meeting* into the communities and practices of the participants.

Take-Home Message

- Navigation of scientific meetings requires research and a seamless execution of a well-formed “preoperative plan” prior to attendance.
- Understanding the components of the meeting and aligning specific areas of interest ahead of time facilitate the attendee to efficiently utilize time and energy.
- Beyond the multitude of educational events, there are many opportunities for networking, equipment demonstrations, and social endeavors.
- It is not necessary, or often not feasible, to do it all.
- At the conclusion of the meeting, it is crucial to dedicate time to synthesize the experience and knowledge gained in order to develop future academic pursuits and collaborations.
- The connections developed at these meetings foster new ideas and relationships between orthopedic surgeons that ultimately improve delivery of patient care and expand research horizons.

- Finally, junior learners are exposed to the specialty outside of traditional lectures and start to build their own network.

Glossary of Meeting-Specific Terms

Scientific meeting Academic gathering of scientific professionals to discuss research and specialty-specific topics.

Networking Forming connections between professionals that are typically used for future collaborations.

Junior learner A young student or physician who attends meetings to expand his/her knowledge and form new relationships.

Roundtable Conversation among experts typically regarding a topic.

Breakout session Short session where a group of attendees discusses a central topic among each other.

Panel discussions A small cohort of experts discussing a topic of interest in front of a large audience.

Symposium A formal discussion among experts in the field regarding a specific topic.

Podium presentation Formal exhibition of a noteworthy research study, typically given in front of an audience.

Instructional course lecture (ICL) An up-to-date educational series, given by a panel of experts, on a specific topic and/or procedure.

Vendor Medical equipment supplier companies present at medical conferences to advertise their product and negotiate partnerships with physicians and medical researchers.

References

1. Borgmann H, DeWitt S, Tsaur I, Haferkamp A, Loeb S. Novel survey disseminated through Twitter supports its utility for networking, disseminating research, advocacy, clinical practice and other professional goals. *Can Urol Assoc J*. 2015;9(9-10):E713–7. <https://doi.org/10.5489/auaj.3014>.
2. Bostic JQ. Practice makes permanent (not necessarily perfect). *J Am Acad Child Adolesc Psychiatry*. 2016;55(9):749–50. <https://doi.org/10.1016/j.jaac.2016.06.007>.

3. Bowman J, Mogensen L, Marsland E, Lannin N. The development, content validity and inter-rater reliability of the SMART-goal evaluation method: a standardised method for evaluating clinical goals. *Aust Occup Ther J*. 2015;62(6):420–7. <https://doi.org/10.1111/1440-1630.12218>.
4. Brick CA, Seely DL, Palermo TM. Association between sleep hygiene and sleep quality in medical students. *Behav Sleep Med*. 2010;8(2):113–21. <https://doi.org/10.1080/15402001003622925>.
5. Devlin R. Value of networking. *Emerg Nurse*. 2016;24(7):17. <https://doi.org/10.7748/en.24.7.17.s23>.
6. Flacco ME, Manzoli L, Boccia S, Capasso L, Aleksovska K, Rosso A, Scaiola G, De Vito C, Siliquini R, Villari P, Ioannidis JP. Head-to-head randomized trials are mostly industry sponsored and almost always favor the industry sponsor. *J Clin Epidemiol*. 2015;68(7):811–20. <https://doi.org/10.1016/j.jclinepi.2014.12.016>.
7. Kay J, Memon M, de Sa D, Duong A, Simunovic N, Athwal GS, Ayeni OR. Five-year publication rate of clinical presentations at the open and closed American shoulder and elbow surgeons annual meeting from 2005-2010. *J Exp Orthop*. 2016;3(1):21. <https://doi.org/10.1186/s40634-016-0059-z>.
8. Kay J, Memon M, de Sa D, Duong A, Simunovic N, Ayeni OR. Does the level of evidence of paper presentations at the Arthroscopy Association of North America annual meetings from 2006-2010 correlate with the 5-year publication rate or the impact factor of the publishing journal? *Arthroscopy*. 2017;33(1):12–8. <https://doi.org/10.1016/j.arthro.2016.05.032>.
9. Kay J, Memon M, Rogozinsky J, de Sa D, Simunovic N, Seil R, Karlsson J, Ayeni OR. The rate of publication of free papers at the 2008 and 2010 European Society of Sports Traumatology Knee Surgery and Arthroscopy Congresses. *J Exp Orthop*. 2017;4(1):15. <https://doi.org/10.1186/s40634-017-0090-8>.
10. Lander ST, Sanders JO, Cook PC, O'Malley NT. Social media in pediatric orthopaedics. *J Pediatr Orthop*. 2017;37(7):e436–9. <https://doi.org/10.1097/BPO.0000000000001032>.
11. Maloney S, Tunnecliff J, Morgan P, Gaida J, Keating J, Clearihan L, Sadasivan S, Ganesh S, Mohanty P, Weiner J, Rivers G, Ilic D. Continuing professional development via social media or conference attendance: a cost analysis. *JMIR Med Educ*. 2017;3(1):e5. <https://doi.org/10.2196/mededu.6357>.
12. Porter ME, Lee TH. From volume to value in health care: the work begins. *JAMA*. 2016;316(10):1047–8. <https://doi.org/10.1001/jama.2016.11698>.
13. Ramnanan CJ, Pound LD. Advances in medical education and practice: student perceptions of the flipped classroom. *Adv Med Educ Pract*. 2017;8:63–73. <https://doi.org/10.2147/AMEP.S109037>.
14. Randviir EP, Illingworth SM, Baker MJ, Cude M, Banks CE. Twittering about research: a case study of the World's first twitter poster competition. *F1000Res*. 2015;4:798. <https://doi.org/10.12688/f1000research.6992.3>.
15. Ritchie A. How to network with fellow physicians to land the job you want. *Med Econ*. 2013. <http://www.medicaleconomics.com/modern-medicine-feature-articles/how-network-fellow-physicians-land-job-you-want>.
16. Robinson R. The art of the elevator pitch: 4 tips for making an impression. *Forbes*. 2017. <https://www.forbes.com/sites/ryanrobinson/2017/09/05/elevator-pitch-tips-making-impression/>.
17. Saperstein AK, Lennon RP, Olsen C, Womble L, Saguil A. Information retention among attendees at a traditional poster presentation session. *Acta Med Acad*. 2016;45(2):180–1. <https://doi.org/10.5644/ama2006-124.178>.
18. Sculco PK, McLawhorn AS, Fehring KA, De Martino I. The future of social media in orthopedic surgery. *Curr Rev Musculoskelet Med*. 2017;10(2):278–9. <https://doi.org/10.1007/s12178-017-9412-9>.
19. Sohn E. Networking: Hello, stranger. *Nature*. 2015;526(7575):729–31.
20. Sumrein BO, Huttunen TT, Launonen AP, Berg HE, Felländer-Tsai L, Mattila VM. Proximal humeral fractures in Sweden—a registry-based study. *Osteoporos Int*. 2017;28(3):901–7. <https://doi.org/10.1007/s00198-016-3808-z>.
21. Torre D, Manca A, Durning S, Janczukowicz J, Taylor D, Cleland J. Learning at large conferences: from the 'sage on the stage' to contemporary models of learning. *Perspect Med Educ*. 2017;6(3):205–8. <https://doi.org/10.1007/s40037-017-0351-3>.
22. Wong LR, Flynn-Evans E, Ruskin KJ. Fatigue risk management: the impact of Anesthesiology Residents' work schedules on job performance and a review of potential countermeasures. *Anesth Analg*. 2018;126(4):1340–8. <https://doi.org/10.1213/ANE.0000000000002548>.



Lukas B. Moser and Michael T. Hirschmann

54.1 Introduction

The first step into unknown territory is always the hardest but the most important one. There are numerous steps in orthopedic residency, where a resident must master unknown areas using unknown methods for unknown problems. It is a steady pursuit for improvement. One has to understand that even the most exciting and original results may not be accepted for publication in a peer-reviewed journal if the presentation and illustration is of mediocre quality. It is of utmost importance to acquire good scientific writing skills for the researchers' armamentarium.

Writing the first scientific paper appears to be one of these difficult steps to master. No one is a born master. In fact, scientific writing has to do more with endurance and discipline than talent. The good news is that scientific writing can be learned as it follows a well-defined structure and writing process.

Writing a scientific article is not like writing a novel. When writing a good novel, the author ideally is in a creative process and tells a comprehensive story using numerous metaphors. A novel

starts with a central conflict, which is not entirely resolved during the story. It leaves the reader with some loose ends, which creates the necessary tension for the reader. This ambiguity and lack of clarity is one essential part of a novel. In addition, a novel does not have to be linear. It could jump from present to past and backwards. Novels are typically written in active person, either first or third person.

In contrast to novels, accuracy and clarity are the most important pillars of good scientific writing. Accurate, clear, and unambiguous expressions of your findings are crucial. Do not use expressions such as "To the best of our knowledge," "up to," or "approximately." Examples of commonly used phrases and their possible impression to the reader clarify the importance of an accurate writing style.

Furthermore, scientific papers are typically written in passive voice and past tense.

Keep in mind:

No one is a born master! Scientific writing can be learned. Endurance and discipline are crucial to improve your skills!

Scientific article \neq novel

L. B. Moser · M. T. Hirschmann (✉)
Department of Orthopaedic Surgery and
Traumatology, Kantonsspital Baselland (Bruderholz,
Liestal, Laufen), Bruderholz, Switzerland

University of Basel, Basel, Switzerland
e-mail: michael.hirschmann@ksbl.ch,
michael.hirschmann@unibas.ch

Accuracy and clarity are keys for scientific writing!

A scientific article is typically structured as the following:

1. A clear and brief title summarizing the major content and findings.
2. A concise, structured, or unstructured abstract serving as a comprehensive summary of your study.
3. The necessary background information to understand the purpose and hypothesis of the study.
4. A concise but comprehensive description of how the study was conducted.
5. The exact and complete but condensed results obtained.
6. A proper yet concise discussion putting the findings into the context of current literature and future research.
7. A brief but comprehensive conclusion giving the clinical relevance of your results.
8. Tables should significantly add to the article and not duplicate the results.
9. Figures should illustrate your results and methods and only be included if highly valuable for the reader.
10. References need to be up to date and formatted with regards to the instructions for authors [20].

This chapter aims to give guidance for young residents aiming to publish their first scientific articles. It is our purpose to guide you through the process of writing a scientific article in a case-based approach. After having read the chapter, the reader should have a guide to start writing such articles and enjoy the process of scientific writing.

54.2 General Comments

Firstly, one should not be intimidated by the blank piece of paper in front of you. You have to accept that the first draft is not aiming for perfection. You

will need to refine and revise your manuscript several times. This takes a considerable amount of your endurance.

Checklist Structure

- *Title*
- *Abstract*
- *Introduction*
- *Methods*
- *Results*
- *Discussion*
- *Conclusion*

Before getting started, you need to complete a detailed review of the current literature with regards to your topic. When you are not familiar with the topic yourself, which is often the case being a junior researcher, it pays off to thoroughly review the topic of your paper. As a more advanced researcher, the topic and the papers published are often well known, which facilitates writing of the introduction. In addition, it makes you aware of the preexisting studies and similarities to your study. Then you might be able to differentiate your article from this group.

Starting with article writing is difficult, and it might be advisable to start with an outline.

One should not worry about the correct syntax, grammar, or language. No one except you will ever read this draft. After having finished your first draft of the manuscript, put it away for days or weeks and then revise again. Do not feel too attached to your writing, critically challenge the content of each sentence and paragraph, and meticulously revise it. Another recommendation is that you show your paper draft to a good colleague or collaborator not being previously involved in the project. Ideally, this person is neither attached to the topic nor the project itself allowing for an unbiased review.

Writing a scientific article follows a clear structure. It gives you as an author less freedom as to what and how you write it, but this also appears to be a chance for the beginner to excel. Clearly, this helps you in getting started [8].

Generally, a scientific paper is structured following the IMRAD acronym. IMRAD stands for Introduction, Methods, Results, and Discussion [16].

After the preparation of a proper study protocol (see Chap. 8), one should start the writing process with the Introduction and Methods followed by the Results and Discussion. The Introduction and Methods can be written after having finished the study protocol and might need to be adapted later on [16].

First step = detailed review of current literature

Start with a draft! Forget about correct syntax, grammar, or perfect language initially!

Introduction
Methods
Results
And
Discussion

For a clinical paper dealing with a prospective randomized controlled trial, the Consolidated Standards of Reporting Trials (CONSORT) should be used. It is an evidence-based, minimum set of recommendations for reporting randomized trials. It offers a standard way for authors to prepare reports of trial findings, facilitating their complete and transparent reporting, and aiding their critical appraisal and interpretation. The CONSORT Statement consists of a 25-item checklist and a flow diagram. The checklist items focus on reporting how the trial was designed, analyzed, and interpreted. The flow chart shows the screening and enrollment processes [19].

54.3 How to Write a Good Title

Although many authors believe that a good title is found easily, the opposite is true. A good title is brief and concise. However, it should give a brief

but comprehensive summary of your study and most important findings. Letchford et al. investigated the benefits of having a shorter title and found that journals publishing papers with shorter titles are more frequently cited [13].

Jargon or colloquial wording has no place in the title. Do not use abbreviations or acronyms in the title. Never pose a question or use exclamation marks in the title. For an easier identification in search engines, it is recommended to use words indexed as Medical Subject Headings (MeSH) in your title [1].

54.3.1 Practical Case Example

Prospective randomized controlled study investigating patients with stiffness after isolated early versus late ACL reconstruction.

54.3.2 Poor Title

What is the difference between early versus late ACL reconstruction?—A single-center RCT.

Use the CONSORT Statement when performing a prospective randomized controlled trial!

Checklist Title

- *Brief and concise*
- *Comprehensive summary*
- *Important findings*
- *No jargon*
- *No abbreviations*
- *No acronyms*
- *No question*
- *No exclamation marks*

54.3.3 Good Title

Early isolated ACL reconstruction shows increased risk of arthrofibrosis—a prospective randomized controlled study.

54.4 How to Write a Good Abstract

Together with the title, the abstract is often the first and only part read by other researchers. Hence, writing of a good abstract directly influences the penetration and scientific power of your study. It decides if your study is acknowledged and cited by your research colleagues.

However, in practice often only limited time is spent for the preparation of a good abstract. Although the abstract is the last written part of a scientific article, in a last effort, one should take enough time to prepare it.

The abstract should be a comprehensive summary of the scientific article. Therefore, you should carefully check the instructions for authors of your target journal before submitting the abstract.

Most journals require rather structured than unstructured formatting. Words are generally limited to 150–350 words highlighting the importance of being brief and concise [1].

54.4.1 Practical Case Example

Prospective study using 3D-CT investigating the influence of the approach in total knee arthroplasty (TKA) on TKA component rotation.

54.4.2 Poor Abstract

Purpose: TKA is a successful treatment option for end-stage osteoarthritis of the knee. Outcome after TKA is influenced by numerous surgery and patient related factors. The purpose of this study was to investigate which factors influence TKA position.

Methods: This study included 200 patients after TKA using either a parapatellar medial or parapatellar lateral approach with tibial tubercle osteotomy. TKA components' position and the whole leg axis were assessed on 3D reconstructed CT scans (sagittal, coronal, and rotational). Mean values of TKA component position and the whole leg alignment of both groups were compared.

The tibial and femoral components were graded as internally rotated, neutral rotation, and externally rotated.

Checklist Abstract

- *Comprehensive summary*
- *Avoid abbreviations*
- *No passive voice*
- *Sample size if percentages are reported*
- *Effect size with confidence intervals*
- *Abstract can be read independently from the main text*
- *Purpose/introduction/background: What is known? Why is this study needed?*
- *Methods: What did I do?*
- *Results: What did I find?*
- *Discussion: What does it mean?*

Results: Two groups were investigated—patients who underwent a medial parapatellar approach (MPA) and a lateral parapatellar approach using a tibial tubercle osteotomy (LPA). Means of tibial component rotation were 2.7323° ER \pm 6.12323 (MPA) and 7.62323° ER \pm 5.4232 (LPA). Patients of group LPA presented a significantly less internally rotated (LPA, 18.43%; MPA, 48.83%) and more externally rotated (LPA, 52.63%; MPA, 22.83%) tibial component ($p < 0.001$). No significant differences were seen for the femoral component position, tibial valgus/varus, and tibial slope.

Conclusion: Patients of group LPA presented a significantly less internally rotated and more externally rotated tibial component. It appears that a LPA tends to externally rotate the tibial TKA component.

54.4.3 Good Abstract

Purpose: The purpose of this study was to investigate if the type of approach [medial parapatellar approach (MPA) versus lateral parapatellar approach with tibial tubercle osteotomy (LPA)] influences rotation of femoral and/or tibial component in total knee arthroplasty (TKA). It

was the hypothesis that MPA leads to an internally rotated tibial TKA component.

Methods: This study included 200 consecutive patients in whom TKA was performed using either a parapatellar medial ($n = 162$, MPA) or parapatellar lateral approach with tibial tubercle osteotomy ($n = 38$, LPA). All patients underwent clinical follow-up, standardized radiographs, and computed radiography (CT). TKA components' position and the whole leg axis were assessed on 3D reconstructed CT scans (sagittal, coronal, and rotational). Mean values of TKA component position and the whole leg alignment of both groups were compared using a t -test. The tibial component was graded as internally rotated ($<3^\circ$ of external rotation (ER)), neutral rotation (equal or between 3° and 6° of ER), and externally rotated ($>6^\circ$ ER). The femoral component was graded as internally rotated [$>3^\circ$ of internal rotation (IR)], neutral rotation (equal or between -3° IR and 3° of ER), and externally rotated ($>3^\circ$ ER).

Results: There was no significant difference in terms of whole leg axis after TKA between both groups (MPA, 0.2° valgus ± 3.4 ; LPA, 0.0° valgus ± 3.5). Means of tibial component rotation were 2.7° ER ± 6.1 (MPA) and 7.6° ER ± 5.4 (LPA). Patients of group LPA presented a significantly less internally rotated (LPA, 18.4%; MPA, 48.8%) and more externally rotated (LPA, 52.6%; MPA, 22.8%) tibial component ($p < 0.001$). No significant differences were seen for the femoral component position, tibial valgus/varus, and tibial slope.

Conclusion: The type of approach (medial versus lateral) significantly influenced tibial TKA component rotation. It appears that a MPA tends to internally rotate the tibial TKA component and a LPA tends to externally rotate the tibial TKA. The anterior cortex should not be used as landmark for tibial TKA component placement when using the lateral approach with tibial tubercle osteotomy.

54.5 How to Write a Good Introduction

This is the part in which you introduce the topic to your reader. Typically, it consists of one page and guides the reader into the topic of your article. The introduction should answer the two key

questions: What is the paper about? Why is it worth being read and published? Arrange your paper from basic to more complex. You should start with a paragraph giving very basic information on the background of your topic. Imagine the structure of the introduction as a funnel. The background information represents the broadest part at the top, which is narrowing down to the specific information of your research topic [2]. However, this first paragraph should not start from "Adam and Eve" but furthermore jump more directly into the topic covered.

54.5.1 Practical Case Example

Prospective randomized controlled study investigating patients with stiffness after isolated early and late ACL reconstruction.

Checklist Introduction

- *Background information from basic to complex*
- *What is known and what is unknown in this specific topic?*
- *Why is the study needed?*
- *Hypothesis (what did we want to know?)*
- *Study aim (how did we answer the research question?)*

Don't lose the red thread of your introduction!

54.5.2 Poor Introduction

"ACL tears are common injuries. ACL reconstruction is the most frequently performed surgery in orthopedics. There are many surgical methods described. The following ACL grafts can be used. The purpose of this study was..."

54.5.3 Good Introduction

An unsolved problem in patients suffering from ACL insufficiency is the timing of ACL reconstruction. [Then describe what is known about it!]

Table 54.1 What you write in your introduction and what the reader probably understands [12]

Introduction	
What you write...	What the reader understands...
It has been known that... It is well known... To the best of our knowledge...	I haven't been bothered to look up the original reference but...
Of great theoretical and practical importance...	Interesting to me...
While it has not been possible to provide definite answers to these questions... Future studies need to clarify the meaning of these results...	The experiment didn't work out, but I figured I could at least get a publication out of it...

Stiffness is an associated problem with early ACL reconstruction... It was our hypothesis that... [Finally end with "The purpose of this study was..."].

Proceed with explaining what is already known, and describe open questions with regard to this topic. The reader should be able to follow the red thread of your introduction.

Like in a good novel, you should guide the reader from basic to complex and climax with your study hypothesis and purpose of the study undertaken.

The last paragraph should include a clear description of your hypothesis and purpose of your study. This part needs to answer the question why you undertook this study and which open research question should be answered.

Please consider the following tips for writing a good introduction (Table 54.1):

1. Perform a detailed and thorough review of current literature.
2. Do not include unnecessary background information! Do not give well-known information! ("Do not start from Adam and Eve!")
3. Keep it as short as possible and as long as needed. Be brief and concise!
4. Do not overstate the clinical value and importance of your work!
5. Ask yourself what you need to know to understand the research topic and purpose of your study. Try to obtain the reader's perspective.
6. Explain to the reader what is novel in your study, which makes it worth being read or

published. Therefore, you need to perform a thorough literature review.

7. End with a clear hypothesis and purpose of your study. The more clear and easy it is, the better it is. Just one or maximum two study questions are optimal.
8. Clearly separate the major from the minor research question.
9. Leave comparison with other studies for discussion [2, 20].

54.6 How to Write a Good Methods Section

This part of a scientific article could be written after having finished your study protocol (see Chap. 8). However, it might need to be adjusted once the study is finished.

Ask yourself: "Will my introduction sell my paper to readers, reviewers, and editors?"

Checklist Methods Section

- *Study design*
- *Setting and subjects*
- *Data collection*
- *Data analyses*
- *Ethical approval*

The methods section should meticulously describe the study design and ideally serve as an instruction for the readership to redo the study. Consider your research study as a specific dish and your methods section as its recipe. You need to list all ingredients and give a detailed description of the cooking process. Only then the dish can be prepared repetitively with a reproducible result [9].

Try to create a clear story line. The methods section links the introduction with the results. Hence, it should be structured from basic to more complex. Generally, one should start with the study design. Here, the authors need to clarify the type of study done (Table 54.2).

Table 54.2 Possible study types for a scientific article [18]

Study type	
Primary research	
Basic research (Experimental research)	
Theoretical	
Method development	Applied
Analytical measurement procedure	Animal study
Imaging procedure	Cell study
Biometric procedure	Genetic engineering
Test development Assessment procedure	Gene sequencing
	Biochemistry
	Material development
	Genetic studies
Clinical research	
Experimental (Interventional)	
Clinical Study	
Phase I study	
Phase II study	
Phase III study	
Phase IV study	
Observational (Noninterventional)	
Therapy study	
Prognostic study	
Diagnostic study	
Observational study with drugs	
Secondary data analysis	
Case series	
Single case report	
Epidemiological research	
Experimental (Interventional)	
Intervention study	
Field study	
Group study	
Biometric procedure	
Test development Assessment procedure	
Observational (Noninterventional)	
Cohort study	
Prospective	
Historical	
Case control study	
Cross-sectional study	
Monitoring, Surveillance	
Description with registry data	
Meta-analysis	
Review	
Systematic	
Simple (narrative)	
Secondary research	

Then, you need to proceed with a clear description of your study sample. The study sample description should include the exact number of patients or subjects included. With regards to patients' basic demographics mean age \pm standard deviation, gender, body mass index, alignment information and other important variables should be given here [9].

54.6.1 Practical Case Example 1

Prospective study using 3D-CT investigating the influence of the approach in total knee arthroplasty (TKA) on TKA component rotation.

Poor description of study sample: A consecutive series of patients who underwent a computed tomography (CT) after primary TKA as clinical routine follow-up or because of knee pain in a university-affiliated hospital were analyzed. Patients were divided into two groups with regard to the used surgical approach.

Good description of study sample: A consecutive series of 200 patients who underwent computed tomography (CT) after primary TKA from 2013 to 2016 as clinical routine follow-up or because of knee pain in a university-affiliated hospital were prospectively collected and retrospectively analyzed. Patients were divided into two groups with regard to the used surgical approach. Group A (lateral peripatellar approach (LPA)) included 38 patients (male/female = 14:24; 67.5 ± 10.4 years), and Group B (medial parapatellar approach (MPA)) included 162 patients (male/female = 56:106; 67.2 ± 9.8 years).

A clear flow chart should allow the reader to understand how many patients were screened, how many were excluded, and how many finally included for this study. It is important to give exact numbers here. Along with an exact description of inclusion and exclusion criteria, the article needs to allow the reader to judge a possible bias in patient selection.

54.6.2 Practical Case Example 2

Prospective study using 3D-CT investigating the influence of the approach in total knee arthroplasty (TKA) on TKA component rotation.

Poor description of inclusion and exclusion criteria: A consecutive series of patients who underwent a computed tomography (CT) after primary TKA as clinical routine follow-up or because of knee pain in a university-affiliated hospital were analyzed. Patients were divided into two groups with regard to the used surgical approach.

Good description of inclusion and exclusion criteria: A consecutive series of 200 patients who underwent a computed tomography (CT) after primary TKA from 2013 to 2016 as clinical routine follow-up or because of knee pain in a university-affiliated hospital were prospectively collected and retrospectively analyzed. Indication for TKA was end-stage osteoarthritis. Only primary TKA were included. Exclusion criteria were any history of infection, tumor. A team of two senior surgeons performed the surgeries using either cruciate retaining or posterior stabilized TKA. The decision to perform a CR or PS TKA was based on the integrity of the posterior cruciate ligament and was done independently from alignment (varus versus valgus knees). Patients were divided into two groups with regard to the used surgical approach.

As a general rule, this part should be sufficiently detailed so that an independent researcher could reproduce the results and thereby validate the study findings [20].

A common error of many authors is to describe the study sample in the results section, but it clearly belongs to material and methods and should be given here. This is also true for review papers [9].

The description of the study sample is followed by a detailed description of tests and experiments done for experimental studies and description of outcome instruments used for clinical studies. In the case a novel methodology is applied, a more detailed description is required. If standard methods are applied such as well-established outcome instruments, it is only necessary to refer to these [20].

For all measurements done, inter- and intra-observer reliability needs to be tested and presented. All measurements, in particular measurements on any image such as radiographs, CT, MRI, or other imaging modalities, should be

done by at least two independent blinded observers twice with an interval of 6 weeks [7].

Inter- and intra-observer reliability as well as accuracy values need to be presented. This could be done, for example, as intra-class correlation coefficients (ICCs), kappa values, or Bland-Altman plots [21].

A statement that ethical approval was obtained from the local ethical committee or institutional review board approval should be included. An increasing number of journals require a document showing ethical approval to be included in your submission. For clinical studies also state that informed consent was obtained by each patient or subject in the study [20].

Some journals require a statement that the study was done in agreement with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards [17].

Describe your study sample as detailed as possible!

Standard methods should be referred to!

Ethical approval is required for every submission in a peer-reviewed journal!

The final paragraph should consist of a proper and complete description of the statistical methodology used [20].

First you have to state which and how the data was presented. For example, “Continuous variables were described using means, standard deviations, and medians. Categorical variables were tabulated with absolute and relative frequencies.”

Another relevant part of the statistical methodology is the exact description of tests used. It is important to differentiate parametric from non-parametric tests. To use parametric tests, the data needs to be tested for normality. Please also

Table 54.3 What you write in your methods section and what the reader probably understands [12]

Methods	What the reader understands...
What you write...	What the reader understands...
Three of the samples were chosen for detailed study...	The results on the others didn't make sense...
Accidentally strained during mounting...	Dropped on the floor...
Handled with extreme care throughout the experiment...	Not dropped on the floor...

report if this was done and the result. Every statistical test used needs to be mentioned here. The level of statistical significance needs to be reported as *p*-value. Generally, it is considered to be $p < 0.05$ [6].

A sample size calculation needs to be presented in all clinical studies. The sample size is an estimation of the number of subjects required to detect a significant difference in a study. If the sample size is too small, then a true difference might turn out nonsignificant although being significant. If the sample size is too large, this would mean that you unnecessarily waste scientific resources, and it might also make even small difference significant (Table 54.3) [5].

Only describe methods which are also presented in the results!

Report your statistical tests used, and explain which test was used to compare which measure with what!

Clinical studies need a sample size calculation!

54.7 How to Write a Good Results Section

This section should be brief and concise. Results are the only thing shown here. It should not contain any description of methodology. Do

Table 54.4 What you write in your results section and what the reader probably understands [12]

Results	
What you write...	What the reader understands...
Typical results are shown...	The best results are shown...
Agreement with the predicted curve is: Excellent Good Satisfactory Fair	Fair Poor Doubtful Imaginary

not interpret your results; simply describe what you have found. Please organize your results in the same logical order and structure as previously reported in material and methods. This makes it better accessible for the reader (Table 54.4) [10].

54.7.1 Practical Case Example

Prospective study investigating the outcome and bone tracer uptake (BTU) in SPECT/CT (single-photon emission computerized tomography in combination with conventional CT) after high tibial osteotomy (HTO) due to symptomatic varus malalignment.

It was the hypothesis that BTU after HTO decreases in the medial compartment and clinical outcome and the degree of correction correlates with BTU and asymptomatic patients after HTO reveals a significantly decreased BTU in the medial subchondral bone.

Twenty-two consecutive patients with 23 knees undergoing medial opening-wedge HTO for medial compartment overloading were assessed pre- and postoperatively (12 and/or 24 months) using Tc-99 m-HDP-SPECT/CT including our 4D-SPECT/CT protocol. BTU was quantified and localized to specific biomechanically relevant joint areas. Maximum absolute and relative values (mean \pm standard deviation, median, and range) for each area were recorded. Pre- and postoperative mechanical alignments were measured. At 24 months after HTO, the WOMAC score was used.

Checklist Result

Brief and concise

Past tense

Presentation without interpretation

Match this section with methods section

Use table and figures to highlight findings

Poor results section: A decrease of BTU in the medial subchondral zones after HTO was found. BTU normalized in all asymptomatic patients. A decrease of BTU was partly seen in the lateral compartments, but the decrease was significantly higher in the de-loaded medial tibial and femoral joint compartment. The achieved average valgus correction of the tibiofemoral angle by HTO was $5.9^\circ \pm 2.8$. There were no adverse events such as pseudoarthrosis, infection, loss of correction, or skin necrosis. The mean WOMAC score pain was 6.2 ± 5.6 , WOMAC stiffness was 2.8 ± 2.4 , and the WOMAC daily activities (0–68) was 17.4 ± 16.4 . The mean total score was 25.4 ± 22.00 after HTO. Less stiffness with regard to the WOMAC score correlated significantly with SPECT/CT BTU. Higher postoperative bone tracer uptake significantly correlated with more pain. Interestingly, no statistical significant associations between SPECT/CT BTU and alignment correction by HTO were found.

Good results section: A significant decrease of BTU in the medial subchondral zones after HTO was found from preoperatively to 12 and 24 months follow-up ($p < 0.01$). BTU normalized in all asymptomatic patients within 24 months. The normalized grading of BTU in SPECT/CT for each anatomical area of the localization scheme is presented in Table 1 (values represent the difference between preoperative and postoperative measurements). A decrease of BTU was partly seen in the lateral compartments, but the decrease was significantly higher in the de-loaded medial tibial and femoral joint compartment ($p < 0.0001$, Fig. 4). The achieved average valgus correction of the tibiofemoral angle by HTO was $5.9^\circ \pm 2.8$. The mean WOMAC score pain (0–20) was 6.2 ± 5.6 , WOMAC stiffness (0–8) was

2.8 ± 2.4 , and the WOMAC daily activities (0–68) was 17.4 ± 16.4 . The mean total score (0–96) was 25.4 ± 22.00 after HTO (Fig. 5). Less stiffness with regard to the WOMAC score correlated significantly with a higher decrease in SPECT/CT BTU ($p < 0.05$).

Higher postoperative bone tracer uptake significantly correlated with more pain ($p < 0.05$). A Spearman correlation analysis revealed no statistical significant associations between SPECT/CT BTU and alignment correction by HTO.

It should be clear that all data is reported and not only the data supporting your hypothesis. Do not report data which has not been mentioned in material and methods. No referencing is allowed in the results section.

Sometimes it appears difficult to decide what exactly is considered a result. For example, if inter- and intra-observer variability is tested for measurements on radiographs for validation of your measurement method, however, it is not the main study question but the radiological results, then the measurements of inter- and intra-observer reliability should be reported not in results but in material and methods.

Figure, tables, and graphs might help to make it better understandable to the reader. Do not duplicate results in text and figures, tables, or graphs. Often not much supporting text is needed here [10].

54.8 How to Write a Good Discussion

The discussion part should put your results into the context of the current literature. In contrast to the previous parts of a scientific article, here the findings need to be explained, interpreted, and debated.

The key questions to be answered are: What is similar and what is different to other studies published? How do your findings help to answer the study question posed? Generally, the discussion should start with a sentence such as “The most important findings of the present study were...”.

However, a pure repetition of your results should be avoided. In the further course of the discussion, the current evidence with regard to the study question needs to be discussed. One common pitfall of the discussion part is to present a review of a considerable number of published studies lacking the interpretation of these papers with regard

Checklist Discussion

- *Summary of main findings*
- *Comparison and interpretation with current literature*
- *Strength and limitations*

to your study question.

Finally, the strength and more importantly the limitations of your study need to be discussed in detail [3].

54.8.1 Practical Case Example

One year clinical and MR imaging outcome after partial synthetic meniscal replacement in stabilized knees using a collagen meniscus implant.

Poor discussion: The purpose of the present study was to evaluate the clinical and radiological outcomes of patients who underwent a medial or lateral collagen meniscus implantation. We hypothesized that good functional results with maintained sport capacity and activity level could be achieved. We further hypothesized that MRI would show no changes in size and signal intensity of the CMI over time. The median preinjury Tegner score was 7 (range 2–10); it decreased preoperatively to 3 (range 0–123). At 1-year follow-up, the median Tegner score was 6 (range 2–10). The mean Lysholm score before surgery was 68 ± 20 and 93 ± 9 at 1-year follow-up. The mean flexion and extension ± 125 standard deviation at 1-year follow-up was $140^\circ \pm 5^\circ$ and $5^\circ \pm 1^\circ$, respectively. Meniscal substitution with the collagen meniscal implant showed excellent clinical 1-year results in a highly active patient group. Significant pain relief and functional

improvement throughout all scores were noted at a minimum of 1-year follow-up. The collagen meniscus implant undergoes significant remodeling, degradation, and extrusion in most of the patients. No difference in outcomes between the medial and lateral CMI was observed.

Good discussion (shortened): The most important findings of this study were twofold: Firstly, the meniscal substitution with the collagen meniscal implant (CMI) showed excellent clinical 1-year results in this large patient series in which most patients also underwent ACL reconstruction. Significant pain relief and functional improvement throughout all scores were noted at a minimum of 1-year follow-up.

A variety of studies have been performed investigating the early to longer-term clinical experience using CMI [1–8]. In agreement with the present study, most of the authors reported that mean Tegner activity and Lysholm scores as well as pain values significantly improved from pre- to postoperatively [5, 6, 9]. There was no significant clinical difference whether a medial or lateral CMI was implanted. Monllau et al. investigated 25 non-consecutive patients at minimum 10 years after CMI implantation [2]. They included almost the same number of patients undergoing an ACL reconstruction and found significantly improved Lysholm and VAS pain scores at 1-year follow-up.

Secondly, with regard to the MRI results, the CMI undergoes significant remodeling, degradation, and extrusion. The new meniscus tissue was well integrated. The size of the meniscus tissue was reduced when compared to the normal meniscus which could be only partially explained by compressive joint loading forces. However, these findings are in agreement with Monllau et al. who found less meniscus volume than expected at minimum 10 years after CMI implantation [2].

One major limitation of this study is the lack of a control group. This study only presents the clinical and radiological outcomes of a consecutive series of patients undergoing collagen meniscus implantation in stable and unstable knees. However, it is one of the biggest case series of patients showing clinical and MRI results 1 year after CMI.

Table 54.5 What you write in your discussion and what the reader probably understands [12]

Discussion	
What you write...	What the reader understands...
It is generally believed that...	A couple of other fellows think so too.
Correct within an order of magnitude	Wrong!
It is clear that much additional work will be required before a complete understanding...	I don't understand it! The results did not turn out to be good enough to make us understand the problem
Future studies are needed...	

Another weakness is that the patients included also underwent a variety of concomitant surgeries. In this study, the majority of patients underwent CMI due to prophylactic reasons, which currently could be considered a dubious indication. This fact might have influenced the results of the study, although this was not shown statistically (Table 54.5).

54.9 How to Write a Good Conclusion

The conclusion should be not more than one or two sentences long and provide the reader with a summary of results and discussion. In particular, the clinical implications of the findings should be highlighted. Please explain why the study is significant to the research world. This part should provide the key message you want to convey with regard to your study. In addition, it is also possible to provide an outlook of future research direction. However, do not just write further investigation is needed or do not announce future studies that might not be completed [3].

54.9.1 Practical Case Example

A prospective, longitudinal, single-cohort study investigating the correlation of depression, control beliefs, anxiety, and a variety of other psychological factors with outcomes of patients undergoing total knee arthroplasty (TKA) in 104 consecutive patients.

Checklist Conclusion

- 1–2 sentences
- Summary of results and discussion
- Clinical Implications
- Why is my study significant?

Poor conclusion: Self-efficacy did not influence clinical scores. More depressed patients showed higher pre- and postoperative WOMAC scores, but no difference in amelioration.

Good conclusion: Depression, anxiety, a tendency to somatize, and psychological distress were identified as significant predictors for poorer clinical outcomes before and/or after TKA. Standardized preoperative screening and subsequent treatment should become part of the preoperative work-up in orthopedic practice.

should be referred to in the text and numbered in order of their citation. Avoid any disagreement of table data and information given in the text [11]!

54.10.1 Practical Case Example

(See Tables 54.6 and 54.7).

54.10 Tips and Tricks for Tables

Tables allow to present a great amount of data in a listed way. It is in particular beneficial for large datasets, which can hardly be described in narrative form. It helps to keep the results section concise and brief. Ask a colleague of yours to check if your table is self-explanatory. Tables

54.11 Tips and Tricks for Figures

Always consider illustrating your article with high-quality figures, which include photographs or drawings. When including any figures, you should ask yourself the question if this figure really adds to the article. Often one is too attached to photographs taken to have an objective view on this matter. It might be beneficial to get a different independent perspective.

The number of figures accepted is dependent from the journal. You will find the actual limits along with resolution and image quality requirements in instructions for authors of each journal. Often figures need to be recut or enlarged to highlight the important part of the figure used. For such purpose, arrows can also be used.

Table 54.6 Bad example of a table included in a scientific article. Clinical outcome scoring preoperatively, 6 weeks, 4 months and 1 year after surgery

Clinical scoring	Pre-operative	<i>p</i>	6 weeks p.o.	<i>p</i>	4 months p.o.	<i>p</i>	1-year p.o.	Overall <i>p</i>
WOMAC pain	52 ± 24 49	***	25 ± 14 22	***	14 ± 16 9	n.s.	13 ± 17 6	***
WOMAC stiffness	54 ± 26 50	***	36 ± 20 30	***	22 ± 20 15	n.s.	20 ± 20 15	***
WOMAC function	48 ± 20 44	***	27 ± 15 22	***	16 ± 15 12	n.s.	16 ± 18 9	***

n.s., *p* > 0.05. **p* < 0.05, ***p* < 0.01, ****p* < 0.001

Table 54.7 Good example of a table included in a scientific article. Clinical outcome scoring preoperatively, 6 weeks, 4 months and 1 year after surgery

Clinical scoring	Preoperative			6 weeks p.o.			4 months p.o.			1-year p.o.			Overall <i>p</i>
	M ± SD	Md	<i>p</i>	M ± SD	Md	<i>p</i>	M ± SD	Md	<i>p</i>	M ± SD	Md		
WOMAC pain	52 ± 24	49	***	25 ± 14	22	***	14 ± 16	9	n.s.	13 ± 17	6	***	
WOMAC stiffness	54 ± 26	50	***	36 ± 20	30	***	22 ± 20	15	n.s.	20 ± 20	15	***	
WOMAC function	48 ± 20	44	***	27 ± 15	22	***	16 ± 15	12	n.s.	16 ± 18	9	***	

Mean ± standard deviation (M ± SD), Median (Md), *p* between measuring points (Wilcoxon-test), and overall *p* (Friedman-test)

n.s., *p* > 0.05, **p* < 0.05, ***p* < 0.01, ****p* < 0.001

Finally, it is also important to refer to the image within your text [11].

Checklist Tables

- *Title reflects content*
- *Title at the top*
- *Self-explanatory*
- *Clear and easy to read*
- *Use correct order*
- *Check if you have followed the instruction for author*

Checklist Figures

- *Title reflects content*
- *Title at the bottom*
- *Use high-quality figures*
- *Does the figure add to the article?*
- *How many figures are accepted by the journal?*
- *Which file format is accepted?*
- *Are instructions for authors followed?*

54.11.1—Practical Case Example

(See Figs. 54.1, 54.2, 54.3, 54.4, 54.5, 54.6, 54.7 and 54.8).

54.12 Tips and Tricks for Reference Section

Formatting of references needs to meticulously follow the individual instructions for authors of each journal. Formatting errors do not increase the confidence of the reviewer in the quality and diligence of your work. Some journals even rapidly reject your work without the possibility of later resubmission. Hence, it is important to follow exactly the instructions for authors. When using reference manager software such as EndNote, Reference Manager, or Mendeley, it is nevertheless important to check before submission that all references are formatted correctly.

One should consider the following principles:

Fig. 54.1 Bad example of a figure included in a scientific article

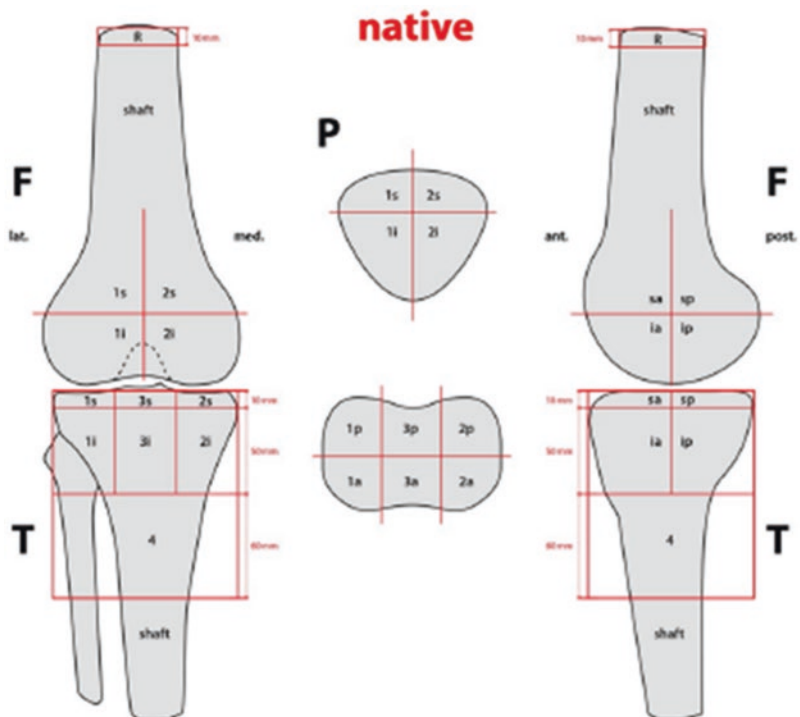


Fig. 54.2 Good example of a figure included in a scientific article [15]

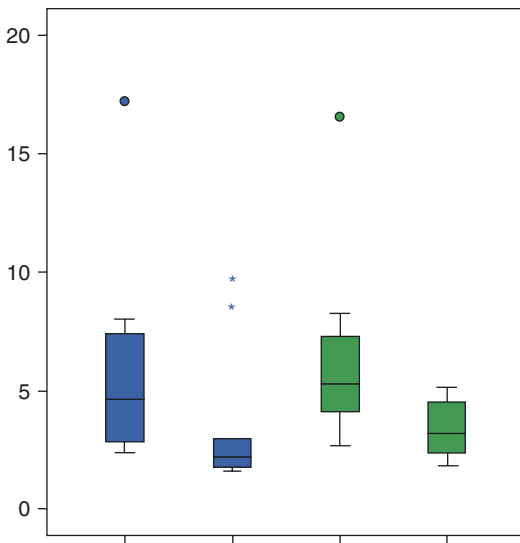
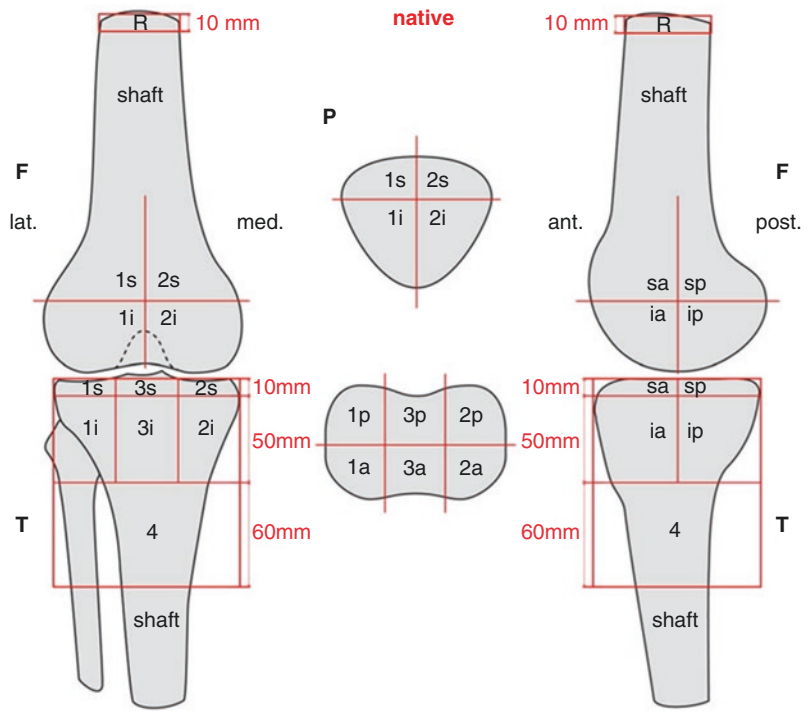


Fig. 54.3 Bad example of a figure included in a scientific article

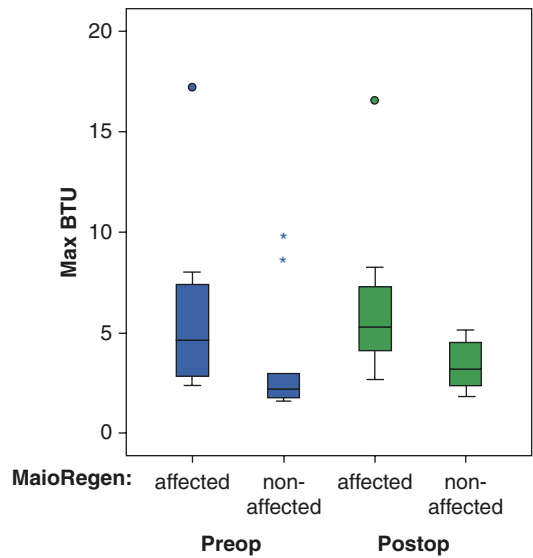


Fig. 54.4 Good example of a figure included in a scientific article [15]

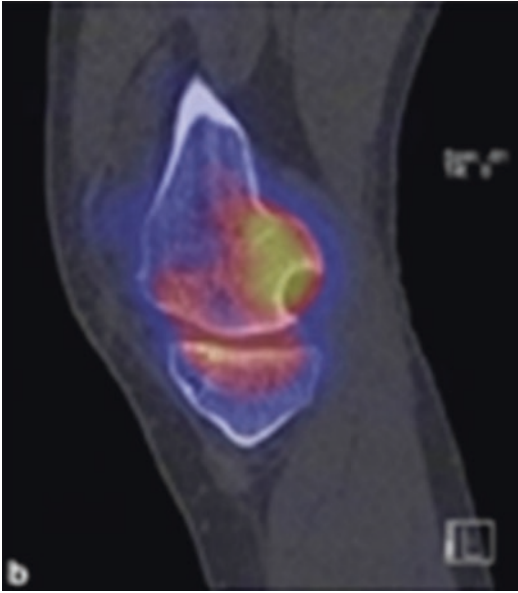


Fig. 54.5 Bad example of a figure included in a scientific article

1. Only use references for which you have read the entire article. Please do not just read the abstract.
2. Try to cite as less references as possible. If a reference is only cited once it must be exceptional. Otherwise delete it.
3. Always cite the original article making the statement you want to reference.
4. Check for duplicates and delete.
5. Please provide a reference for every statement you make.
6. Please avoid overly self-referencing. Only cite your own paper if these really contribute here. Some authors tend to cite as many of their own papers as possible.
7. Use the latest and updated references for your article. Recheck just before submission of article.

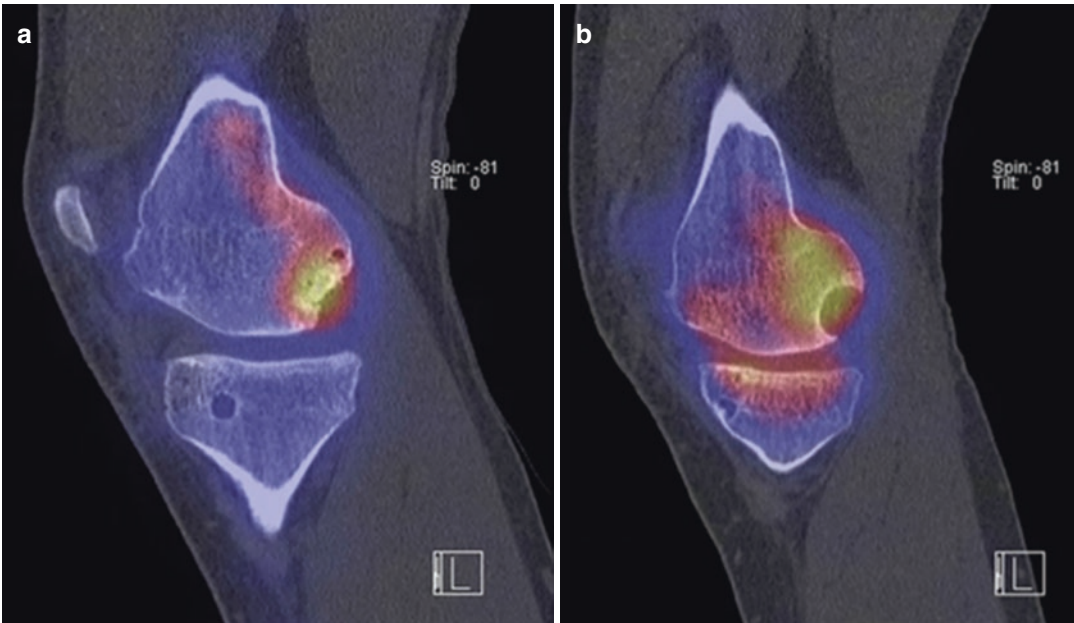


Fig. 54.6 Good example of a figure included in a scientific article [15]

Fig. 54.7 Bad example of a figure included in a scientific article

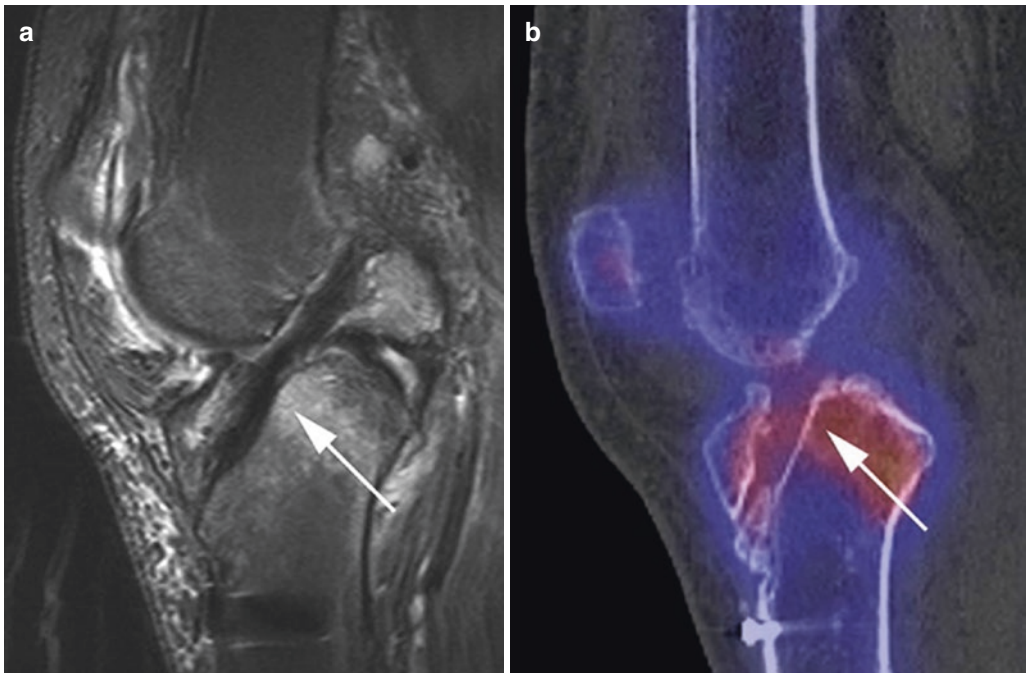
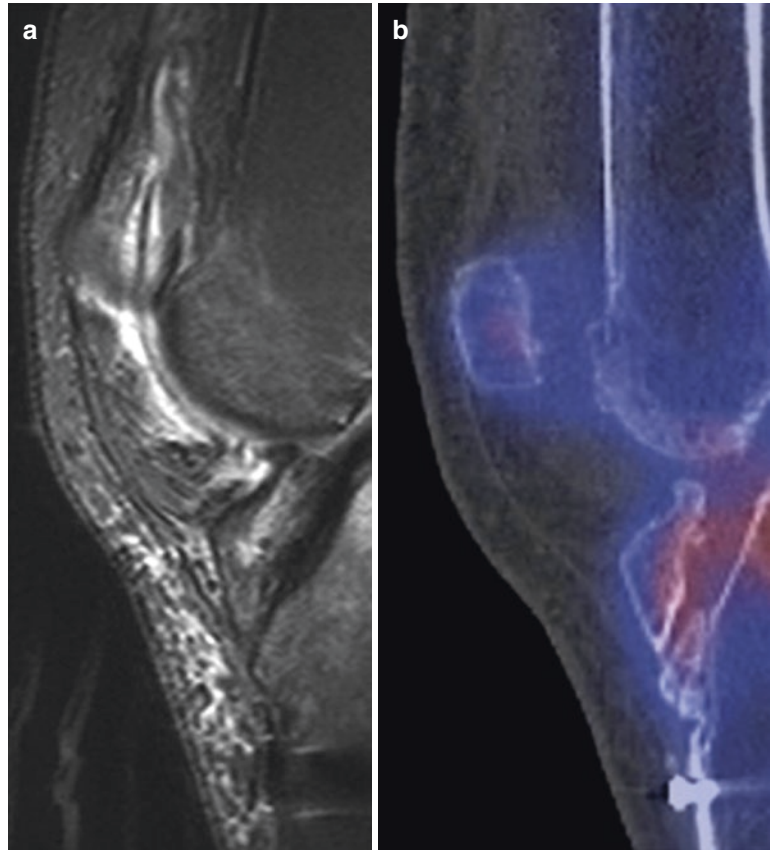


Fig. 54.8 Good example of a figure included in a scientific article [14]

8. Try to cite relevant papers of the target journal. Editors appreciate it as an interest in their journal, and it might improve citation scores [4].

Checklist References

- Use reference manager software
- Use the requested output style
- Cite the original reference
- Recheck the final reference list

Take-Home Message

- Writing the first scientific paper appears to be a difficult step to master. Nevertheless, scientific writing follows a well-defined structure. Your first step is to complete a detailed review of current literature with regard to your topic. Afterward it is advisable to start with an outline. Now you can benefit from your previously written comprehensive study protocol.
- Accurate, clear, and unambiguous expression of findings is crucial and improves the quality of your manuscript. Even though you have exciting and original results, your manuscript may not be accepted for publication in a peer-reviewed journal if the presentation and illustration are of mediocre quality. Therefore, it is of utmost importance to acquire good scientific writing skills for your research armamentarium.
- Finally, you need to be certain that you meticulously follow the instructions stipulated by your target journal prior to submission.

References

1. Cals JW, Kotz D. Effective writing and publishing scientific papers, part II: title and abstract. *J Clin Epidemiol.* 2013;66:585.
2. Cals JW, Kotz D. Effective writing and publishing scientific papers, part III: introduction. *J Clin Epidemiol.* 2013;66:702.
3. Cals JW, Kotz D. Effective writing and publishing scientific papers, part VI: discussion. *J Clin Epidemiol.* 2013;66:1064.
4. Cals JW, Kotz D. Effective writing and publishing scientific papers, part VIII: references. *J Clin Epidemiol.* 2013;66:1198.
5. Guller U, Oertli D. Sample size matters: a guide for surgeons. *World J Surg.* 2005;29:601–5.
6. Harris JD, Brand JC, Cote MP, Faucett SC, Dhawan A. Research pearls: the significance of statistics and perils of pooling. Part I: clinical versus statistical significance. *Arthroscopy.* 2017;33:1102–12.
7. Hassink G, Testa EA, Leumann A, Hugle T, Rasch H, Hirschmann MT. Intra- and inter-observer reliability of a new standardized diagnostic method using SPECT/CT in patients with osteochondral lesions of the ankle joint. *BMC Med Imaging.* 2016;16:67.
8. Kotz D, Cals JW. Effective writing and publishing scientific papers—part I: how to get started. *J Clin Epidemiol.* 2013;66:397.
9. Kotz D, Cals JW. Effective writing and publishing scientific papers, part IV: methods. *J Clin Epidemiol.* 2013;66:817.
10. Kotz D, Cals JW. Effective writing and publishing scientific papers, part V: results. *J Clin Epidemiol.* 2013;66:945.
11. Kotz D, Cals JW. Effective writing and publishing scientific papers, part VII: tables and figures. *J Clin Epidemiol.* 2013;66:1197.
12. Lacroix JR. A key to scientific research literature. *Can Med Assoc J.* 1971;104:1080.
13. Letchford A, Moat HS, Preis T. The advantage of short paper titles. *R Soc Open Sci.* 2015;2:150266.
14. Mathis DT, Hirschmann A, Falkowski AL, Kiekara T, Amsler F, Rasch H, et al. Increased bone tracer uptake in symptomatic patients with ACL graft insufficiency: a correlation of MRI and SPECT/CT findings. *Knee Surg Sports Traumatol Arthrosc.* 2018;26(2):563–73. <https://doi.org/10.1007/s00167-017-4588-5>.
15. Mathis DT, Kaelin R, Rasch H, Arnold MP, Hirschmann MT. Good clinical results but moderate osseointegration and defect filling of a cell-free multi-layered nano-composite scaffold for treatment of osteochondral lesions of the knee. *Knee Surg Sports Traumatol Arthrosc.* 2018;26(4):1273–80. <https://doi.org/10.1007/s00167-017-4638-z>.
16. Peh WC, Ng KH. Basic structure and types of scientific papers. *Singap Med J.* 2008;49:522–5.
17. Rickham PP. Human experimentation. Code of ethics of the World Medical Association. Declaration of Helsinki. *Br Med J.* 1964;2:177.
18. Rohrig B, du Prel JB, Wachtlin D, Blettner M. Types of study in medical research: part 3 of a series on evaluation of scientific publications. *Dtsch Arztebl Int.* 2009;106:262–8.
19. Schulz KF, Altman DG, Moher D, CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomized trials. *Open Med.* 2010;4:e60–8.
20. Vitse CL, Poland GA. Writing a scientific paper—a brief guide for new investigators. *Vaccine.* 2017;35:722–8.
21. Watson PF, Petrie A. Method agreement analysis: a review of correct methodology. *Theriogenology.* 2010;73:1167–79.



Common Mistakes in Manuscript Writing and How to Avoid Them

55

Eleonor Svantesson, Eric Hamrin Senorski, Kristian Samuelsson, and Jón Karlsson

55.1 Introduction

Conducting a study of high quality is challenging and requires both effort and discipline. It is a process that may take years. However, when the day finally arrives, the day when you have the results, you are of course eager to make them official and let them impact current practice and evidence-based medicine.

There is just one thing; the manuscript needs to be written. Considering all the work and time you have invested in conducting the study, you will likely feel ambitious about presenting your work in the best possible way and to get it published in a high-impact journal [1]. On the other

hand, many researchers feel that final drafting of the manuscript, before crossing the finish line, feels more or less like climbing a mountain. Writing is something that may not come natural for some researchers and clinicians. However, with discipline and some straight-forward tools for writing, it may not be that complicated. And, when you start to master the art, you might find that it was not as hard as you thought from the beginning. And, in the end, you may even find writing a manuscript fun.

55.1.1 Dedication

The main key for success is to be dedicated to your work. If you really want to become a successful researcher, you need to be passionate about your research topic and be prepared to invest time and effort in your work. A high motivation and an ability to put your goals in front of you can help you defeat even the hardest struggles. To be dedicated also means that you are willing to learn and are able to acknowledge your shortcomings. Many researchers before you have experienced exactly the same struggle that you might be feeling. You should therefore view every challenge as an opportunity to learn from these experienced researchers. It is time to create new ground. You are now in a position of writing the manuscripts which will form the textbooks used by your future colleagues. Mentors that can share experiences and tips about how to prepare a

E. Svantesson (✉)

Department of Orthopaedics, Institute of Clinical Sciences, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

E. H. Senorski

Department of Health and Rehabilitation, Institute of Neuroscience and Physiology, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

K. Samuelsson · J. Karlsson

Department of Orthopaedics, Institute of Clinical Sciences, The Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden

Department of Orthopaedics, Sahlgrenska University Hospital, Mölndal, Sweden
e-mail: jon.karlsson@telia.com

manuscript are a highly valuable asset when aiming to develop a good skill for writing. Nevertheless, it is up to you whether you have an open mind and use this opportunity wisely.

55.2 Common Mistakes

Even though your research may be of good quality, there are some common mistakes that might increase the risk of your manuscript ending up in the “rejection box” instead of being published. Interestingly, some of these mistakes may sound obvious, but from an editor’s experience, the below listed mistakes keep getting repeated over and over again:

1. *The manuscript is too comprehensive.* A manuscript should be short and concise. This is an area where a too high motivation to publish the manuscript actually may be your downfall. Considering all your effort in conducting the study, it is understandable if you would like to present every aspect of it, including all the data, all previous topic-related literature and discuss all possible findings of your study. In other words, you might be tempted to write everything you know. However, ask yourself—what was really the main purpose and hypothesis of this study? What is new and what are the most important findings of this study? The production rate of research articles today is extremely high, and to include too many results in a single article may cause the most important results to drown in text and data. Moreover, an excessive length of the manuscript may entail that fewer individuals will take time to sit down and read and reflect over the findings. Therefore, choose your focus points and stick to them. Sometimes less is more. In fact, most articles are too long. Most articles repeat information that is already well known, and this is hardly ever necessary. It has been said that “...a manuscript should be as long as necessary, but as short as possible...” and this is true. A manuscript should never be so long that it is boring to read.
2. *There is no clear line of argument.* An article should be enjoyable to read. The flow of your writing is crucial for this purpose. You need to decide for yourself before writing what your line of argument is and arrange your argument(s) in a logical flow. Logical flow will increase the readability of the manuscript and the chances of getting it published.
3. *Unnecessary repetitions and statements.* The vast number of published articles entails that you could theoretically repeat tangentially related results and well-established facts to an eternity. Again, aim to have your manuscript focused and concise. There is no need to repeat your own findings or findings from topic-related literature all over again. Sometimes it is necessary to assume that the reader should already be aware of some basic knowledge of the area and instead leave room for the readers who are interested to make a deepened review of the literature themselves. Thus, use the opportunity to refer to other studies wisely so that the reader could find further information if wanted, without presenting the results of each study in the reference list in detail.
4. *Instructions to authors.* A practical key is to read and follow “Instructions to authors”. The time it takes to read them is always well invested. Way too often, it is obvious that authors have not read the instructions. Another mistake, which is very annoying to editors, is to resubmit a manuscript that has been rejected by another journal without answering the raised comments, changing the format, or bothering to look at the different instructions for the particular journal you have now chosen. Fact Box 55.1 summarizes common mistakes in manuscript writing.

Each manuscript is comprised of several essential sections that are worth a thorough review in order to present each section in the best possible way [2, 5]. Systematically writing each section of the manuscript usually facilitates writing since it allows the author to feel how the manuscript successively takes form under constant critical review. Therefore, let’s give some focus on each one of these sections.

Fact Box 55.1: Common Mistakes in Manuscript Writing

- The manuscript is too long and intends to cover a too wide range of topics
- There is no clear line of argument and the text is not organized in a logical flow
- Similar findings and statements are getting repeated over and over again
- The manuscript goes into details of basic knowledge instead of using the opportunity to refer to studies where the reader independently could find more information if necessary
- The authors have not taken the time for reading the journal guidelines thoroughly
- Not answering reviewers' comments in case of resubmission to a different journal

55.2.1 Title

The title should capture the reader's interest immediately. The title should be as short as possible and give the reader an idea of the study and the main results. A common mistake is that the title is too neutral and only mirrors the area of investigation. Instead, let the title speak. Let it be loud and clear and shout out a direct finding of your study. The title should be a statement and never a question.

55.2.2 Abstract

In general, the abstract follows the same main structure for all journals; however, there may be some slightly different subheadings and word limits between journals. An abstract should include the study purpose, a brief presentation of the methodology, the main results, and a conclusion. Remember, the reader must be able to understand which material and methods have been applied in order to understand the results that are presented in the abstract. This means that details of the methodology can be limited in the abstract, but the full description should be included in the main text of the manuscript. The abstract also functions as an opportunity to raise

interest in the study. Therefore, take some time to formulate the conclusion as direct as possible and, preferably, somewhat controversial. The ultimate goal of the abstract is to make the reader curious to find out more about how you have reached your conclusion and wanting to read the full-text article. Preferably, the conclusion of the abstract should be the same as in the text. Many clinical journals ask for level of evidence. This information should then be added at the end of the abstract if required.

55.2.3 Introduction

As the word implies, the introduction should introduce the reader to the topic of the study. However, some authors write a far too comprehensive introduction that may, paradoxically, instead assuage the interest of the study before even reaching the results. Another common mistake is to start discussing results in the introduction. The introduction should be used to raise some interesting discussion topics and highlight relevant questions regarding the topic. These questions will then be answered and discussed in the discussion section. The introduction should end by turning focus toward your study. A clear purpose and a hypothesis for your study should be stated at this point. These are essential elements that the reader will bear in mind when reading all the following sections of your manuscript. A good rule of thumb is that the introduction should be no more than one manuscript page in length.

55.2.4 Materials and Methods

You have probably heard it before, but it is worth repeating: The methods should be so well described that the reader should be able to repeat your study like a recipe from a cookbook without trouble. This means that the methods section should be detailed, clear, and honest. The text should have a good flow and a readable language. Instead of trying to create a literary masterpiece

of the methods section, keep it simple and precise. The most important aspects to focus on are the inclusion and the exclusion criteria, the description of the intervention or the experiment, and a clear presentation of the outcome measurements. Consider using flow diagrams or tables to illustrate your test setup. There is a number of reporting guidelines published that can help you structure your methods section depending on the type of study. Examples of such guidelines are the CONSORT [4] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement [3]. The statistical analysis should be described under a separate subheading, where all calculations should be clearly reported. This may include power, sample size, sensitivity analysis, and drop-out analysis. In fact, sample size calculation is always necessary in order to ensure that the statistical power is adequate. Statistics are really a separate science in itself and do therefore not hesitate to ask for professional help in order to ensure that the statistical analysis is correctly presented. Two common mistakes are to leave out information about the sample size calculation and the IRB approval. It can hardly be stated enough that (almost) all studies need an IRB approval.

55.2.5 Results

Now you have reached the part where you finally are allowed to present the results of your study. The most important thing of this section is that the results are presented objectively. Preferably, you have prepared a thorough study protocol before conducting your trial where you already have prepared an outline for your results section. There should be no subjective influence in the result section whatsoever; save that for the discussion part. Take time to “get to know the data” and to decide how you best present it. There is no need to present all the findings in the text; choose the most important one(s) to present in detail in the text. The text is in turn complemented by tables and figures in order to display all data. Aim for writing a short results section where the results are not duplicated in text and in tables or figures. A good rule of thumb is that the results

section should not be longer than one manuscript page.

55.2.6 Discussion

The way you start the discussion is important. In a few initial sentences, you should preferably summarize the most important findings of your study, which function as a foundation for the rest of your discussion. The discussion should be written based on your results, and these should be compared and contrasted with previous research, not the other way around.

Thus, the discussion is not a forum for a general discussion and presentation of other studies. Primarily ask yourself: What did your study show? Thereafter, discuss these findings in relation to previous findings. Is likely that your results are true? What supports your findings compared with other studies and what findings of your study might be contradicted based on previous research? Also, focus on the clinical relevance of your findings in the discussion. This is especially important if you have conducted pre-clinical research and are aiming for a clinically oriented journal. Finally, all researchers know that no study is perfect. There are always limitations and confounding factors that could influence a result of a study. To be honest and humble about such, potential factors increase the trustworthiness of a study. Furthermore, an understanding of limitations of a study will generate new ideas for future studies and encourage honest research. Therefore, think through your study limitations, and clearly present and discuss them at the end of the section. All too often, limitations are not as well reported as they should be.

55.2.7 Conclusion

The conclusion should be based on statistically significant findings from your study and nothing else. There is room for some slight speculation; however, such speculations should mainly be included in the discussion and never in the conclusion. The conclusion should be a brief, true, and concrete statement of the evidence that

Fact Box 55.2: A General Outline of a Manuscript and the Contents of Each Section

Introduction	Raise some interesting discussion topics and highlight relevant questions regarding the topic. State a clear purpose and a hypothesis. Keep it short, approximately one manuscript page
Materials and methods	The reader should be able to repeat your study by reading this section. Keep it simple and precise. Consider it a cookbook
Results	Present the results objectively. Focus on presenting the details for the most important findings; present data in table and figures to minimize the length of this section
Discussion	Start by summarizing the most important findings. Compare and contrast your results with previous research
Conclusion	Should be a brief, true, and concrete statement of the evidence that your study has contributed to. It is not an extended discussion

your study has contributed to and nothing else. In certain journals, there is also room for a brief comment of the clinical relevance of your study, especially if you have conducted an experimental study. A common mistake is that the conclusion is an extended discussion. Fact Box 55.2 summarizes the general manuscript outline.

55.2.8 References

One of the easiest and most straight-forward aspects of preparing a manuscript should be to get the references correct. Nevertheless, this is an area that is all too often defective among submitted manuscripts. The most common explanation for this is probably that the authors simply have not been thorough when reading the journal guidelines before writing the reference section of the manuscript. The most common mistakes are

that the references are either in an incorrect format or that they are not up to date. Each journal has specific guidelines for how to prepare the references regarding the order and the format. Read the guidelines carefully—this is always well-invested time—and get to know your reference system so that you can adjust the references accordingly. To avoid submitting a manuscript with references that are not up to date, update your references just before submitting your manuscript to the journal. This is also logical, as you might have started the study a couple years back and much could have happened during this time. Taken together, the two common mistakes are format errors and the use of non-updated references.

55.2.9 Figures and Tables

Again, this is an area where all journals have different preferences of how these should be submitted and where in the manuscript they should be located. Thus, the primary way to avoid mistakes is to read the guidelines of the journal. Another aspect to consider is that both tables and figures should be designed in a way that makes them understandable, independently of the rest of the article. It is important that figures and tables are accompanied by descriptive legends that are self-explanatory. Every figure should be able to be read as “stand-alone.” This includes, for example, a presentation of abbreviations and key ideas. When used correctly, tables and figures are valuable methods for presenting large volume of data, to visualize the results and to keep the result text section short. Avoiding repetition is an important part of tables and figures. They should give the details of the results, but not repeat them. Another mistake is the figure quality; figures should be drawn by professional medical artists and not by amateurs.

Take-Home Message

- To write a manuscript is a process that takes time and effort.
- Aim for a short and concise manuscript, with a clear line of argument.
- Read the journal guidelines thoroughly and follow them in detail.

- Mentors that can share experiences and tips about how to prepare a manuscript are a highly valuable asset when aiming to develop a good skill for writing.

References

1. Katz MJ. From research to manuscript: a guide to scientific writing. New York: Springer; 2009.
2. MacArthur CA, Graham S, Fitzgerald J. Handbook of writing research. New York: Guilford; 2006.
3. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol*. 2009;62(10):1006–12. <https://doi.org/10.1016/j.jclinepi.2009.06.005>.
4. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332. <https://doi.org/10.1136/bmj.c332>.
5. Swales JM, Feak CB. Academic writing for graduate students: essential tasks and skills. Ann Arbor: The University of Michigan Press; 2012.