



The PCORI Methodology Report

PCORI Methodology Committee

David Hickam, Annette Totten, Alfred Berg, Katherine Rader, Steven Goodman,
Robin Newhouse, *Editors*

Revised Draft for Board Acceptance

November 2013

Editorial Team for Methodology Stories and Examples

Writers/Editors: Ayodola Anise, Eric Johnson, Zachary Meisel, Edwin Reid, Lauren Saxton

Designer of Review Materials: Lauren Saxton

Chief Editor for Stories: Mark Helfand

Acknowledgements

PCORI wishes to thank staff who assisted with creating this document: Joe Selby, Julie Miller, Amy Grossman, Blake Whitney, Julie McCormack, William Silberg, and Bryan Luce. Former PCORI Methodology Committee chair Sherine Gabriel and former vice chair Sharon-Lise Normand provided leadership in development of the draft PCORI Methodology Report and the final PCORI Methodology Standards. Additional contributors are noted in Appendix F.

PCORI is solely responsible for the final content of this report.

Suggested citation for this report:

PCORI (Patient-Centered Outcomes Research Institute) Methodology Committee. 2013. "The PCORI Methodology Report." pcori.org/research-we-support/research-methodology-standards.

About PCORI

PCORI was authorized by the Patient Protection and Affordable Care Act of 2010 as a non-profit, nongovernmental organization and is charged with helping patients, clinicians, purchasers, and policy makers make better-informed health decisions by “advancing the quality and relevance of evidence about how to prevent, diagnose, treat, monitor, and manage diseases, disorders, and other health conditions.” It does this by producing and promoting high-integrity, evidence-based information that comes from research guided by patients, caregivers, and the broader healthcare community.

PCORI is committed to transparency and a rigorous stakeholder-driven process that emphasizes patient engagement. PCORI uses a variety of forums and other mechanisms to obtain public input to enhance its work.

Patient-Centered Outcomes Research Institute

1828 L St., NW, Suite 900

Washington, DC 20036

Phone: (202) 827-7700

Fax: (202) 355-9558

Email: info@pcori.org

Follow us on Twitter: @PCORI

The PCORI Methodology Report

PCORI Methodology Committee

David Hickam, Annette Totten, Alfred Berg, Katherine Rader, Steven Goodman, Robin Newhouse, *Editors*

Revised Draft for Board Acceptance

November 2013

PCORI Methodology Committee

Robin Newhouse (Chair), Professor and Chair, Organizational Systems and Adult Health, University of Maryland School of Nursing

Steven Goodman (Vice Chair), Associate Dean for Clinical and Translational Research, Professor of Medicine & Health Research and Policy, Stanford University School of Medicine

Naomi Aronson, Executive Director of the Blue Cross and Blue Shield Association Technology Evaluation Center

Ethan Basch, Director of the Cancer Outcomes Research Program, University of North Carolina, Chapel Hill

Alfred Berg, Professor, Department of Family Medicine, University of Washington School of Medicine

David Flum, Professor, Department of Surgery; Adjunct Professor, Health Services and Pharmacy, University of Washington School of Medicine

Mark Helfand, Staff Physician, Portland VA Medical Center; Professor of Medicine and of Medical Informatics and Clinical Epidemiology, Oregon Health & Science University

Michael Lauer, Director, Division of Cardiovascular Sciences, National Heart, Lung, and Blood Institute

David Meltzer, Chief, Section of Hospital Medicine; Director, Center for Health and the Social Sciences; Chair, Committee on Clinical and Translational Science; Associate Professor, Department of Medicine, Department of Economics, and the Harris School of Public Policy Studies, University of Chicago

Brian Mittman, Director, VA Center for Implementation Practice and Research Support, Department of Veterans Affairs, Greater Los Angeles VA Healthcare System

Sebastian Schneeweiss, Associate Professor of Medicine and Epidemiology, Harvard Medical School; Vice Chief, Division of Pharmacoepidemiology and Pharmacoeconomics, Brigham and Women's Hospital

Jean Slutsky, Director, Center for Outcomes and Evidence, Agency for Healthcare Research and Quality

Mary Tinetti, Gladdys Phillips Crofoot Professor of Medicine, Epidemiology, and Public Health, Division of Geriatrics, Yale University School of Medicine

Clyde Yancy, Chief, Cardiology, Northwestern University Feinberg School of Medicine; Associate Director, The Bluhm Cardiovascular Institute, Northwestern Memorial Hospital

EXECUTIVE SUMMARY

Introduction

The Patient-Centered Outcomes Research Institute (PCORI) was authorized by the Patient Protection and Affordable Care Act (PPACA) of 2010. The institute's mission was established to help people make better-informed healthcare decisions and improve healthcare delivery and outcomes by producing and promoting high-integrity, evidence-based information that comes from research guided by patients, caregivers, and the broader health care community. It does this by advancing comparative clinical effectiveness research focused on providing useful information about the outcomes that are important to patients and those who care for them. This approach, patient-centered outcomes research (PCOR), provides information designed to help people make informed healthcare decisions. It allows their voices to be heard in assessing the value of the healthcare options available to them.

The PCORI Methodology Committee provides critical guidance to the institute in advancing this mission, and to the research field more broadly. The committee was established by the PPACA to "develop and improve the science and methods of comparative clinical effectiveness research." This report summarizes the committee's work to date in meeting that charge. It is a substantially revised version of a draft report that PCORI made available for public comment in July 2012.

All research programs must adhere to best practices in the planning, design, and conduct of every individual research project. Such practices can be implemented through specific guidelines that are accepted by stakeholders. To enhance the mission of providing high-quality information through the best possible research, the PCORI Methodology Committee has developed guidelines for PCORI itself and for the broad community of researchers who conduct PCOR. The first component is a clinical prioritization process and research translation framework that helps make PCORI's research agenda transparent and accessible to stakeholders. These concepts inform the creation of translation tables that facilitate PCORI's efforts to implement its priorities and make the most efficient use of scarce resources. The second component of the guidelines is a set of standards that provide specific guidance for the design and conduct of individual PCOR projects. The PCORI Methodology Standards underwent extensive public review and revision, and they have been formally adopted by the PCORI Board of Governors.

Research Prioritization and Translation Framework

Research prioritization is a coordinated set of activities used for identifying the important clinical information gaps that can be addressed by new PCOR. These activities offer an approach to align a program of PCOR with questions that patients and clinicians have about what works best, for whom, and under what circumstances. PCORI is striving to ensure that there is a good match between its research priorities and patients' and clinicians' information needs. PCORI research needs to be directed toward providing the answers patients need for health decisions. PCORI's framework for prioritizing research topics includes the following factors:

- Disease incidence, prevalence, and burden (with emphasis on chronic conditions);
- Gaps in evidence in terms of clinical outcomes, practice variation, and health disparities;
- Potential for new evidence to improve health, well-being, and quality of care;
- Effect on national expenditures associated with healthcare treatment, strategy, or health conditions;
- Patient needs, outcomes, and preferences; and
- Relevance to patients and clinicians in making informed health decisions.

Broad public input is an essential component of implementing a PCOR agenda. However, obtaining the perspectives of patients, caregivers, and other stakeholders is useful only if their input can be evaluated meaningfully. Value-of-information (VOI) analysis is a technique for identifying those research questions that have the greatest potential to improve population health. VOI accounts for uncertainty in the health benefits and risks associated with alternative treatment choices, the ability of research findings to alter that uncertainty, and the resulting care decisions (Raiffa and Schlaifer, 1961; Meltzer, 2001). After research questions have been prioritized—and researchers have had the opportunity to propose specific projects—peer and stakeholder review is the final stage in selecting projects and moving them toward completion. The review process identifies those proposals most likely to fulfill PCORI's objectives and agenda. PCORI has particular advantages and responsibilities in developing its approach to peer review. Incorporating patients and other stakeholders in peer-review activities is a novel approach, but it also makes the process more complex. To maintain integrity, separation between those under review and those reviewing proposals must be safeguarded.

Even when the need for new PCOR has been clearly identified and prioritized, the quality of evidence provided by new research depends on the design and quality of the studies conducted. The choice of study designs has practical implications for the timeliness, validity, and relevance of the research agenda that PCORI pursues. The translation table guides the choice of study designs by helping balance such factors as validity of the resulting evidence, appropriate use of scarce research resources, and timeliness of results for specific questions. When research designs clearly match the questions patients and other stakeholders consider important, research results should be more readily accepted and implemented. The translation table's purpose is to provide guidance for two main tasks: 1) choosing a basic study design; and 2) determining additional design details. The translation table is not a fixed product; it needs to be developed based on the information available for each general research question. And table development is based on a framework that summarizes the important factors underlying the tradeoffs among alternative study designs. The translation framework comprises the following principles:

- Keep the research question and the methodology separate.
- Focus on clarifying tradeoffs.
- Place individual research studies in the context of a research program.
- Have the choice of study design take into account state-of-the-art research methodology.

PCORI Methodology Standards

The PCORI Methodology Standards are specific recommendations for researchers that designate the minimal requirements for following PCOR best practices. The PCORI Methodology Committee developed the standards by following a systematic process. The committee surveyed the range of potential standards, narrowed its scope to those it deemed most important, solicited feedback through a public comment period, revised the draft standards, and confirmed the final set of standards through consensus of its members.

Building on the work of the Institute of Medicine (IOM 2011), the committee started with the following definition of a standard:

- A process, action, or procedure for performing PCOR that is deemed essential to producing scientifically valid, transparent, and reproducible results. A standard may be supported by scientific evidence. When such evidence is unavailable, a standard may be endorsed by

reasonable expectation that the standard helps achieve the desired level of quality in PCOR or by broad acceptance of the practice in PCOR.

- The research practices recommended by the standard can be feasibly implemented.

Initial topics were chosen to reflect areas where there were either substantial deficiencies or inconsistencies in how available methods were applied in practice or where there was specialized knowledge in how best to conduct research or to surmount barriers to the effective dissemination of research results (Helfand et al. 2011; Lohr. 2007; Schneeweiss et al. 2012). Methodological standards aim to do this by improving the way each research question is selected, formulated, and addressed, and how the findings are reported. Standards can also help prevent the use of flawed methods to answer research questions. The current PCORI Methodology Standards are a first installment of what will be an ongoing process of both broadening the scope of the standards and revising existing ones. PCORI uses the standards in its peer review of applications that investigators submit to PCORI for research funding.

Characteristics of the Methodology Standards

This set of standards does not represent a complete, comprehensive set of all requirements for high-quality PCOR; rather, they address an initial group of topics that are likely to contribute to improvement in PCOR quality and value. Specifically, the standards focus on selected methodologies and issues that reflect areas where there are either substantial deficiencies or inconsistencies in how available methods are applied in practice or where there is evidence supporting the recommended practices. Most of the standards can be considered “minimal,” meaning that they are necessary for sound science but should not inhibit further evolution of methods. Other standards are designed to promote transparency: how properly to communicate—in both study protocols and published reports—exactly what was planned and what was done. The standards are based on scientific justification, either from empirical studies, when this is available, or from theoretical work about research methods. Because PCOR can use a variety of specific designs, approaches, and techniques, the PCORI Methodology Standards cannot address all possible issues in clinical research.

PCORI's 47 standards fall into 11 categories, the first five of which are relevant to most PCOR studies. Researchers should refer to all of these cross-cutting standards when planning and conducting their projects. These categories are:

- Formulating research questions
- Patient-centeredness
- Data integrity and rigorous analyses
- Preventing and handling missing data
- Heterogeneity of treatment effect (HTE)

The other six categories of standards are applicable to particular study designs and methods. Two of the categories provide guidance on developing specific types of data and using them in studies:

- Data registries
- Data networks as research-facilitating infrastructures

The final four categories apply to studies that have varying designs and purposes. The standards in each of these categories should be used for guidance when it is relevant to a particular study:

- Causal inference methods
- Adaptive and Bayesian trial designs
- Studies of diagnostic tests
- Systematic reviews

The PCORI Methodology Standards are listed by title in the table at the end of this executive summary. The full text of the standards can be found in **Appendix A: PCORI Methodology Standards**.

Departures from basic good research practices are partially responsible for mismatches between the quality and relevance of the information research provides and the information patients need to make informed clinical decisions. One of the most important components of this foundation is a commitment to transparency in research.

Transparency enables stakeholders and researchers to verify research findings. Many of the standards promote transparency by requiring such best practices as:

- Asking a well-formulated research question;
- Preparing detailed research protocols; and
- Adhering to guidelines for registering studies and reporting results.

This first set of PCORI Methodology Standards establishes a foundation for maintaining best PCOR practices. However, PCORI expects that these standards will change and expand over time so that they address the full spectrum of PCOR inquiries and approaches. PCORI is also interested in advancing the science of patient-centered study design, patient and stakeholder engagement, dissemination, and implementation.

PCORI is pursuing a comprehensive, coordinated approach to promote the uptake of these standards. This approach includes engaging all stakeholders who might use the standards, collaborating with existing entities and initiatives to strengthen research practices and to facilitate use of the standards, and creating reporting and surveillance mechanisms. Future activities might include developing training resources, checklists, and other tools to support researchers' decisions and practices, as well as checklists and other decision-support tools for peer reviewers.

Cross-Cutting Standards for PCOR

1: Standards for Formulating Research Questions

- RQ-1** Identify gaps in evidence
- RQ-2** Develop a formal study protocol
- RQ-3** Identify specific populations and health decision(s) affected by the research
- RQ-4** Identify and assess participant subgroups
- RQ-5** Select appropriate interventions and comparators
- RQ-6** Measure outcomes that people representing the population of interest notice and care about

2: Standards Associated with Patient-Centeredness

- PC-1** Engage people representing the population of interest and other relevant stakeholders in ways that are appropriate and necessary in a given research context
- PC-2** Identify, select, recruit, and retain study participants representative of the spectrum of the population of interest and ensure that data are collected thoroughly and systematically from all study participants
- PC-3** Use patient-reported outcomes when patients or people at risk of a condition are the best source of information
- PC-4** Support dissemination and implementation of study results

3: Standards for Data Integrity and Rigorous Analyses

- IR-1** Assess data source adequacy
- IR-2** Describe data linkage plans, if applicable
- IR-3** A priori, specify plans for data analysis that correspond to major aims
- IR-4** Document validated scales and tests
- IR-5** Use sensitivity analyses to determine the impact of key assumptions
- IR-6** Provide sufficient information in reports to allow for assessments of the study's internal and external validity

4: Standards for Preventing and Handling Missing Data

- MD-1** Describe methods to prevent and monitor missing data
- MD-2** Describe statistical methods to handle missing data
- MD-3** Use validated methods to deal with missing data that properly account for statistical uncertainty due to missingness
- MD-4** Record and report all reasons for dropout and missing data, and account for all patients in report
- MD-5** Examine sensitivity of inferences to missing data methods and assumptions, and incorporate into interpretation

5: Standards for Heterogeneity of Treatment Effects

- HT-1** State the goals of HTE analyses
- HT-2** For all HTE analyses, pre-specify the analysis plan; for hypothesis-driven HTE analyses, pre-specify hypotheses and supporting evidence base.
- HT-3** All HTE claims must be based on appropriate statistical contrasts among groups being compared, such as interaction tests or estimates of differences in treatment effect.
- HT-4** For any HTE analysis, report all pre-specified analyses and, at minimum, the number of post-hoc analyses, including all subgroups and outcomes analyzed

Standards for Specific Study Designs and Methods

6: Standards for Data Registries

- DR-1** Requirements for the design and features of registries
- DR-2** Standards for selection and use of registries
- DR-3** Robust analysis of confounding factors

7: Standards for Data Networks as Research-Facilitating Structures

- DN-1** Requirements for the design and features of data networks
- DN-2** Selection and use of data networks

8: Standards for Causal Inference

- CI-1** Define analysis population using covariate histories
- CI-2** Describe population that gave rise to the effect estimate(s)
- CI-3** Precisely define the timing of the outcome assessment relative to the initiation and duration of exposure
- CI-4** Measure confounders before start of exposure and report data on confounders with study results
- CI-5** Report the assumptions underlying the construction of propensity scores and the comparability of the resulting groups in terms of the balance of covariates and overlap
- CI-6** Assess the validity of the instrumental variable (i.e. how the assumptions are met) and report the balance of covariates in the groups created by the instrumental variable for all instrumental variable analyses

9: Standards for Adaptive and Bayesian Trials

- AT-1** Specify planned adaptations and primary analysis
- AT-2** Evaluate statistical properties of adaptive design
- AT-3** Specify structure and analysis plan for Bayesian adaptive randomized clinical trial designs
- AT-4** Ensure clinical trial infrastructure is adequate to support planned adaptation(s)
- AT-5** Use the CONSORT statement, with modifications, to report adaptive randomized clinical trials

10: Standards for Studies of Diagnostic Tests

- DT-1** Specify clinical context and key elements of diagnostic test study design
- DT-2** Study design should be informed by investigations of the clinical context of testing
- DT-3** Assess the effect of factors known to affect diagnostic performance and outcomes
- DT-4** Structured reporting of diagnostic comparative effectiveness study results
- DT-5** Focus studies of diagnostic tests on patient-centered outcomes, using rigorous study designs with preference for randomized controlled trials

11: Standards for Systematic Reviews

- SR-1** Adopt the Institute of Medicine (IOM) standards for systematic reviews of comparative effectiveness research, with some qualifications.

Contents

Executive Summary	i
Introduction	1
Section I: Patient-Centered Outcomes Research	5
Section II: Prioritizing Research Questions and Development of the Translation Table	11
Setting Priorities for Patient-Centered Outcomes Research	
Choosing a Study Design: Translation Framework	
Section III: PCORI Methodology Standards	21
Introduction	
Cross-Cutting Standards for PCOR	
Standards for Specific Study Designs and Methods	
Section IV: The Context for Implementing the Methodology Standards and Next Steps	75
Methodology Stories and Examples	
CER Wins	
Two Studies Using Hospitals to Improve Care	6
A Surprise Finding That Led to Immediate Changes in Treatment for Abnormal Heart Rhythms	9
Including Greater Varieties of Patients in Studies Proves Valuable	10
Patient Voices	
PCORI Reviewers	15
Lucinda Shore	30
Juli	31
A Woman with Fibromyalgia	32
Sarah	40
Suzanne	51
Research in Practice	
Chest Pain Choices	7
Analyzing the Value of Information	12
PCORI Prioritization Pilot	14
Pamela Williams	30
Missing Data	39
Data Registries	49
Research Stories	
Bias in Last Observation Carried Forward Method	41
Heterogeneity of Treatment Effects	44

National Joint Registry of England and Wales	48
Human Immunodeficiency Virus	61
Getting off the Ventilator	71
Aspirin for the Prevention of Colorectal Cancer	72

Appendices

A: PCORI Methodology Standards	A-1
B: Response to Public Comments	B-1
C: Recommended Actions and Research Recommendations	C-1
D: The PCORI Methodology Committee’s Approach to the Authorizing Legislation	D-1
E: Translation Framework	E-1
F: References	F-1
G: Contributors	G-1

This page intentionally left blank.

INTRODUCTION

The Patient-Centered Outcomes Research Institute (PCORI) was authorized by the Patient Protection and Affordable Care Act (PPACA) of 2010. The institute was established to help people make informed healthcare decisions and improve healthcare delivery and outcomes by producing comparative effectiveness research (CER) that is guided by patients, caregivers, and the broader healthcare community. According to the National Academy of Sciences' Institute of Medicine (IOM), CER "compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve the delivery of care" (IOM 2009). PCORI has developed a program of patient-centered outcomes research (PCOR) that meets this goal by emphasizing research that examines choices and clinical outcomes that are meaningful to patients. This research, which takes into account patients' values and preferences, helps individuals and their caregivers make informed healthcare decisions.

The [federal legislation](#)¹ that authorized PCORI required that its research program be based on rigorous scientific methods. Specifically, PCORI was directed to pursue two early activities that would help to support its scientific mission. The first is developing methodology standards that "provide specific criteria for internal validity, generalizability, feasibility, and timeliness of research and for health outcomes measures, risk adjustment, and other relevant aspects of research and assessment with respect to the design of research." The second is creating a translation table that would provide guidance to "determine research methods that are most likely to address each specific research question." PCORI has completed its initial work on these requirements, and this report summarizes the results.

The healthcare system is complex and addresses a broad range of diseases and conditions. Consequently, patients often face many options when seeking medical care. The need for better evidence about clinical effectiveness is great, and PCORI (like all other organizations that support medical research) must choose among many possible research opportunities. Prioritizing topics and determining which research designs can provide information that is both useful and timely require a systematic process, which will be addressed in Section II of this report. The translation framework that forms the basis of building translation tables is a component of this activity and will also be described in that section.

¹ Available at pcori.org/assets/PCORI-Authorizing-Legislation-032310.pdf

Methodological standards offer guidance for conducting high-quality PCOR studies. Regardless of the chosen research design, some of the standards are applicable and should be considered when planning and conducting a study. Through a careful process of evaluating needs and obtaining input from a broad range of stakeholders, PCORI has addressed this need by developing a set of methodology standards (PCORI Methodology Committee 2012). The standards are listed in Section III of this report, which also includes the description and rationale for each of the overarching topics that guided the development of the specific standards.

To illustrate the importance of the issues addressed in this report, we have included four sets of stories and examples, each with a different focus:



Although these stories and examples are not intended to describe specific standards or to endorse particular research approaches, they demonstrate that good methodology makes a difference.

Developing a translation framework and the initial set of PCORI Methodology Standards for patient-centered outcomes research is a milestone but not a destination. In its ongoing work, the PCORI Methodology Committee will reconsider, refine, and widen the scope of the standards to include the full spectrum of PCOR questions and approaches (Gabriel and Normand 2012). The pursuit of medical knowledge has evolved from predominantly case studies to clinical trials and observational studies that use increasingly complex designs and analytic methods. Given that future advances are expected in research methodology, PCORI has a commitment to evaluate and update the guidance that it provides to the research community. In developing the initial set of methodology standards, the PCORI Methodology Committee also developed a set of recommended actions to provide a direction for future activities (see **Appendix C: Recommended Actions and Research Recommendations**).

This page intentionally left blank.

SECTION I: PATIENT-CENTERED OUTCOMES RESEARCH

For many illnesses, there are choices among alternative treatments and strategies. The decisions that patients make account for much of their personal experience with health care. Ultimately, health and well-being are the result of living with the consequences of these choices. Some people do their own research at the library or on the Internet; others count on the advice of a trusted friend or healthcare provider. No matter how these decisions are made, they should be based on the best information available. Clinical research is an important source of information to guide the decisions that people make every day about their health and medical care.

Often there is a gap between the information that people need and the information that research provides. Sometimes, research is conducted using people with a limited range of characteristics, such as age, sex, race, and complexity of conditions. Research may be restricted to treatment in sophisticated research centers rather than typical community settings. And research often does not include all the outcomes that most people think are important.

A program of clinical research should provide reliable, relevant, and useful health-related evidence for decision makers, especially patients and their caregivers. PCOR aims to provide information that can help guide the answers to such patient-centered questions as:

1. Given my personal characteristics, conditions, and preferences, what should I expect will happen to me?
2. What are my options, and what are their potential benefits and harms?
3. What can I do to improve the outcomes that are most important to me?
4. How can clinicians and the care-delivery organizations they work in help me make the best decisions about my health and health care? (Examples of how healthcare delivery systems have participated in comparative research can be found in **CER Win: Two Studies Using Hospitals to Improve Care.**)



CER WINS: Two Studies Using Hospitals to Improve Care

Comparative effectiveness research (CER) often examines drugs, medical devices, or other specific treatments. However, it sometimes compares how health systems operate. For example, CER studies have considered strategies that hospitals use to provide consistent treatment. Other studies have compared methods that hospitals use to avoid errors. The studies ask which strategies are most effective to help hospitals provide treatments in the best way.

What Strategies Help Hospitals Avoid Infections?

Too often, patients get infections while in the hospital. Such hospital-acquired infections are common and can be deadly. Each year, 17,000 hospital patients die from hospital-acquired infections. In 2004, for example, 1,000 patients developed serious infections in Michigan hospitals. The rate was similar in other states. But such infections are often preventable.

A major source of the infections is thin tubes, called central line catheters, inserted into large veins. In the Keystone Intensive Care Units (ICU) project, most Michigan hospitals participated in a large, prospective, observational study that examined a new process for preventing hospital-acquired infections. Teams of doctors and nurses followed a series of simple steps for inserting and removing catheters from large veins. The hospitals reminded staff to follow the steps, provided real-time feedback, and implemented other changes (Goeschel and Pronovost 2008) to make safety for patients everyone's job. The team compared Michigan hospitals, which made the changes, with hospitals in nearby states that did not. After two years, among patients 65 years or older, there were no catheter-associated infections in the ICUs at most of the Michigan hospitals, and the Michigan patients had lower death rates than similar patients at the other hospitals (Lipitz-Snyderman et al. 2011; Pronovost et al. 2006).

What This Study Adds: This large study showed the value of a hospital procedure as it was performed throughout many different types of hospitals in Michigan. Therefore, the results will probably apply to communities of patients who seek care in various settings.

Minutes Count: Does a Delay in Treatment Matter for Heart Attack Patients?

During a heart attack, the time it takes to get the patient treatment can matter a great deal. For some patients, delays can lead to serious heart problems and even death.

For certain heart attacks, the best treatment is called angioplasty, a procedure that unblocks a crucial blood vessel. Specialized cardiologists thread a balloon-like device through the patient's blood vessel, then inflate it. Some hospitals are not equipped for this. Patients needing angioplasty often are transferred to hospitals that offer the procedure. Randomized controlled trials have compared patients who were moved and received angioplasty with those treated at the original hospital. When there were no delays, the transferred patients fared better. Rapid transfer, however, isn't always feasible.

How long a delay is too long for a patient to benefit from angioplasty? A recent observational study used large registries of data on patients to answer this question. The study compared ST Elevation Myocardial Infarction patients who were transferred to hospitals that could perform angioplasty versus those who were treated with fibrinolytic therapy at the first hospital. The results demonstrated that delays to reperfusion are common among patients transferred for primary treatment and that the mortality advantage for transfer declines as treatment delays lengthen. When the delay was 120 minutes or longer—which was true for 48 percent of patients in the community—angioplasty offered no benefit over drugs. The benefit of angioplasty occurred in those patients transferred rapidly to angioplasty-capable hospitals (Pinto et al. 2011).

What This Study Adds: By studying a larger, less highly selected group of patients and hospitals, this study expanded the clinical trial results, making clear when a patient who is having a heart attack can benefit from being transferred to another hospital for angioplasty and when it is just as good to get immediate treatment with fibrinolytic therapy. The study also shows that registries—particularly when combined with sophisticated analytic techniques—can play a key role in informing clinical decisions.



RESEARCH IN PRACTICE: Chest Pain Choices

Soon after Annie LeBlanc, her husband Michel Demers, and their children moved from Canada to the United States, Michel began experiencing chest pain. They share their story along with Erik Hess, MD, MS, of the Mayo Clinic and leader of the PCORI-funded Chest Pain Choice study (Hess et al. 2012).

Annie LeBlanc: A few months back, my husband wasn't feeling well at all. He was experiencing chest pain. His father and grandfather had died suddenly of a heart attack, so he was very concerned about this condition. He phoned me at work. We were new in town, and we didn't have many family or friends at the time. We rushed home to find a babysitter for the kids. Then we rushed to the ER. They got so many tests very quickly, but then they came back to us saying that "everything seems to be normal." Still, they wanted to run more tests. We stayed for another two hours. More blood tests, EKG, and chest X-rays.

Michel Demers: We were very worried about what was happening.

LeBlanc: All this time, to be honest, we wanted to get back to the kids. The doctors came back to us saying that everything was all right, but they didn't want to take any chances, so they wanted to admit him for a stress test in the morning. But I was aware of the choices we had. So, I started to ask questions. Instead of options and choices, we got comments such as, "You don't want your husband to be alright?" and "We're pretty sure this is nothing bad, but if this was my brother, I wouldn't let him go home."

I asked the doctor, "What is the risk of heart attack in the next month?"

"It's low."

"How low?"

"Low, but we still want to make sure."

My husband felt worse because he didn't understand and couldn't express himself (he speaks French primarily). Finally, we saw someone who could explain the risk. He knew the results of the clinical comparison studies that showed the difference between staying and going home. He said, "Okay, here are your choices. Your risk is very low. I can keep you under observation and have the stress test in the morning. I can have you seen by a cardiologist within 48 hours. Or you can go to your primary care provider for follow-up."

We didn't have a primary care provider at the time. We chose to follow up with the cardiologist. That was what we wanted and that was what happened. In the end, everything was fine. No stress test done, even as an outpatient. Now we are part of the research team looking at shared decision making in chest pain. What we did at the beginning really was to tell our story. As the researchers think about guiding patients through the experience of making decisions about chest pain, we make sure that it matches what we were experiencing. It was our journey. And they needed to understand it. We were part of every part of every step of the research process. We provided input on the...decision aid. We pointed out what was missing and how it was to be distributed. And then what we were expecting in terms of outcomes that meant something to us. It's amazing. Every time we meet, our experience shapes the way the protocol or intervention is being used.

Erik Hess: One of the things that I was surprised by, as a provider and researcher, is that if we treat low-risk patients automatically the same as the moderate-risk patients, the patients perceive their risk as moderate. Good evidence allows us to communicate the risk in a much clearer way, and then we can mitigate their anxiety by including them in the decision-making process.

LeBlanc: Research gave us choices, the choices reduced my husband's stress, and, I think, made him healthier.

For more information, see *The Chest Pain Choice Decision Aid*, available at <http://circoutcomes.ahajournals.org/content/5/3/251.full>.

PCORI aims to catalyze and promote research that can be used by patients, caregivers, and clinicians to answer these questions, and for evidence to be useful, it must be relevant and readily available. (See **Research in Practice: Chest Pain Choices.**)

PCORI and other research funding agencies have the opportunity to promote research that optimally supports decision making by patients, caregivers, clinicians, and policy makers. To support such decisions, research needs to answer the questions it intends to answer. Promoting rigor in the methods used by researchers can help to make their findings valid and more useful.

An important aim of medical research is to conduct studies that provide accurate estimates of benefit or harm. But there is always some uncertainty. Methodological standards encourage research approaches that allow correct conclusions to be drawn. PCORI's Methodology Standards have been developed to address specific criticisms of some clinical research. While much existing research has successfully avoided these problems, some studies fall short of providing high-quality, useful information. PCORI's Methodology Standards are designed to provide a layer of protection from these problems in future research. A key challenge is getting the questions right. Researchers often choose questions and outcomes that they consider to be interesting and important. Sometimes these are not the questions and endpoints most relevant to people who need information. Researchers sometimes do not focus on outcomes that are difficult, expensive, or take too much time to assess. (For an example where choice of outcome made a difference, see **CER Wins: A Surprise Finding That Led to Immediate Changes in Treatment for Abnormal Heart Rhythms.**)

Another challenge is that research sometimes focuses on patients with a narrow set of characteristics and conditions. Often practical reasons influence this choice: it takes a much larger study to account for differences among patients, and the bigger the study, the greater the cost. Sometimes there are scientific purposes at play: for example, narrowing the number of variables in a trial of a new drug makes it more likely that any effects are due to the drug and not something else. Sometimes researchers want to include a broader range of patients but are unable to do so because they have trouble recruiting study participants who represent the full spectrum of patients. (To learn about two trials taking a wide approach, see **CER Wins: Including Greater Varieties of Patients in Studies Proves Valuable.**)



CER WINS: A Surprise Finding That Led to Immediate Changes in Treatment for Abnormal Heart Rhythms

Patients who survive a heart attack may not be out of danger. In the months after the attack, their lives can be threatened by abnormal heart rhythms. In 1987, researchers examined how well three medicines worked to prevent abnormal heart rhythms. The trial enrolled adults who had suffered a heart attack within the previous two years and later experienced abnormal rhythms. The study tallied heart attacks and deaths for 10 to 18 months. The researchers compared the effects of the medicines and an inactive substance. They found that the drugs did suppress abnormal heart rhythms. But the researchers got a surprise. All three medicines were associated with a higher death rate than the inactive substance. After this finding was reported, physicians stopped prescribing the medicines to heart patients (CAST-II Investigators 1992; Echt et al. 1991).

What This Study Adds: Before this study, it was taken for granted that the drugs would reduce death rates because they were proven to reduce some abnormal rhythms. The medicines were widely prescribed but had not been compared directly. The surprise finding was discovered because the trial measured patient-relevant clinical outcomes (death rates), whereas previous studies only looked at intermediate outcomes (heart rhythm). The trial led to an immediate and lasting change in treatment for patients who had previously had a heart attack.



CER WINS: Including Greater Varieties of Patients in Studies Proves Valuable

Some randomized trials of medical treatments use strict eligibility criteria to select people who are similar to each other. The participants all receive the treatment in the same way in settings that are alike. These similarities make it easier for researchers to show that differences in results come from the treatment being tested rather than other factors. But such carefully controlled trials may not show how a treatment will affect a wide variety of patients in a range of settings. Randomized trials using broad populations, diverse settings, and “simple” eligibility criteria can provide strong results that change medical practice.

Drug Reduces Heart Attack Deaths

One of the first “large simple trials,” called ISIS-1, enrolled 16,000 people in 14 countries. Each person had gone to a hospital with symptoms of a heart attack. Within a few hours, the participants were randomly assigned to one of two groups. One group received standard treatment, which at that time did not include drugs called beta blockers. The participants in the other group had a beta blocker infused into their veins and later took the drug by mouth. Patients treated with the beta blocker had a 15 percent lower death rate in the first week of the study compared with a control group. No significant difference in mortality was noted between the groups after the first week (ISIS-1 1986).

What This Study Adds: This study showed that beta blockers are an effective therapy for nearly all groups of patients who may be having a heart attack. The study changed the way heart attack patients are treated.

Screening for Abdominal Aneurysm

The aorta, the largest blood vessel in the body, sometimes balloons out into what is called an abdominal aneurysm. If this aneurysm ruptures, the internal bleeding can lead to death. A screening with ultrasound can identify an abdominal aneurysm before any symptoms appear. Would such screening of a large group of people be worthwhile? A British trial randomly assigned 68,000 men between ages 65 and 74 to receive—or not receive—an invitation for a screening ultrasound. Over the next seven years, the study found that the men invited to the initial screening had about half as many deaths due to an abdominal aneurysm as those not invited for screening (Kim et al. 2007).

What This Study Adds: By keeping the criteria for entering the study broad and conducting it in the setting of normal clinic practice, investigators strengthened the evidence that the intervention is effective.

SECTION II: PRIORITIZING RESEARCH QUESTIONS AND DEVELOPMENT OF THE TRANSLATION TABLE

Generating new research involves actions by many individuals and organizations in addition to the researchers themselves. Aligning these activities can help to ensure that research programs are optimally efficient in terms of both time and resources. Because PCORI is an organization that provides funding to individual research teams and promotes the development of high-quality PCOR, its methodology work embraces a broad set of initiatives.

Setting Priorities for Patient-Centered Outcomes Research

Establishing a specific research agenda is a core duty of PCORI. Unless there is a good match between research priorities and the information needs of patients and clinicians, methodological standards will have limited effect. PCORI research needs to be directed toward providing the answers patients need for health decisions. Formulating decisions and defining questions for research are complex processes requiring expertise and open-mindedness (Bravata et al. 2005; Buedo 2005), careful formulation of research questions, and a commitment to patient-centeredness.

Research Topic Prioritization

PCORI's Board of Governors is charged with developing, refining, prioritizing, and selecting among research investments. To guide this process, PCORI uses a framework that includes the following factors:

- Disease incidence, prevalence, and burden (with emphasis on chronic conditions);
- Gaps in evidence in terms of clinical outcomes, practice variation, and health disparities;
- Potential for new evidence to improve health, well-being, and the quality of care;
- Effect of health conditions and treatments on national expenditures;
- Patient needs, outcomes, and preferences; and
- Relevance to patients and clinicians in making informed health decisions.

The best way to ensure that these factors are assessed successfully when comparing alternative topics is to obtain broad input from patients and other stakeholders. PCORI also has an obligation to spend its resources effectively and efficiently. When there is more than one acceptable research approach available, the potential added cost of alternative study designs should be balanced against the potential

value and timeliness of the likely research results. Such techniques as value-of-information (VOI) analysis—a statistical method for estimating the average improvement in outcomes that may be expected by obtaining additional information (Meltzer et al. 2011; Claxton and Sculpher 2006)—may be useful in clarifying tradeoffs between study cost and the degree of certainty expected from study results. (See **Research in Practice: Analyzing the Value of Information**.)



RESEARCH IN PRACTICE: Analyzing the Value of Information

In choosing what research to fund, PCORI must balance the cost of a project against the potential usefulness of the information it can produce. Value-of-information (VOI) analysis is a tool for making such choices. A recent study looked into whether VOI analysis would be useful in a process in which healthcare stakeholders help decide which research to fund (Carlson et al. 2013). In this study, the researchers worked with stakeholders who were advising a group that funds trials of cancer treatments. Josh Carlson, MPH, PhD, is an assistant professor at the University of Washington and an affiliate faculty member at the Fred Hutchinson Cancer Research Center, both in Seattle.

How did you explain VOI to the stakeholders in your study?

Josh Carlson: We prepared an educational document on value of information. It was only three pages long. We tried to use simple language to describe VOI. We also gave presentations based on that document and allowed the stakeholders to ask questions and interact with us.

In the educational document, did you use an example to illustrate the concept?

Carlson: One example we used was a drug prescribed for advanced breast cancer. It was approved based on data from a single phase two trial that showed that the drug had an effect on the cancer but did not show that it increased quality or length of life. The Food and Drug Administration approved the drug. But doctors and policy makers were unsure whether they should offer the drug to patients now or wait for additional evidence given the remaining uncertainty.

What did your study show?

In our study, we asked thirteen stakeholders to rank three potential cancer genomic research areas. They indicated their preferences both before and after receiving VOI information. The VOI information appeared to influence stakeholder rankings, with seven changing their ranking. Further, most of the stakeholders reported that they had found the analysis useful in their decision making.

How do you see VOI analysis being integrated into deciding what healthcare research to fund?

Carlson: VOI analysis is useful in that it can help people compare across a range of technologies but can best serve as one factor among multiple decision-making criteria. I think it works best within specific research areas. It gets a bit harder when you ask people to decide between completely different research programs. Ultimately, the goal is to help maximize the impact of research.

There are four key components of research prioritization: topic generation, systematic reviews and gaps analysis, VOI analysis, and peer and stakeholder review. Although these steps tend to be pursued sequentially, the research prioritization process is iterative; for example, results of VOI analysis might influence topic-generation initiatives.

Topic Generation

Topic generation is necessary to ensure that PCORI considers a sufficient number and range of topics before it selects topics for research funding. Including patients and other stakeholders is an important part of this process. Some empirical research, mostly conducted outside of the United States, has shown that patient involvement can produce more relevant research questions and results that are more useful for making decisions (Nass et al. 2012; Oliver et al. 2008; Patient Partner 2012). Without adequate input from patients and other stakeholders, research priorities may not fully reflect patient perspectives on potential benefits or risks. Understanding this, PCORI is testing and developing novel and existing approaches to obtaining patient and other stakeholder input in research topic generation. (See **Research in Practice: PCORI Prioritization Pilot.**)

Systematic Reviews and Gap Analysis

Systematic reviews can identify gaps in knowledge that underlie uncertainty among patients and clinicians. Sometimes systematic reviews generate new questions. For example, a pooled analysis of several studies can reveal an important finding that was not evident in the individual studies. Systematic reviews can also highlight a key question for patients that none of the studies has answered. Identifying gaps in the existing literature and deficiencies in completed studies should reduce investments in research that are unlikely to help answer important questions. Ethics also require that researchers avoid recruiting patients into unneeded studies. Using gap analyses—based on systematic reviews also fosters transparency and accountability in funding prioritization. These concepts have informed one of PCORI's methodology standards that provides guidance for formulating research questions.



RESEARCH IN PRACTICE: PCORI Prioritization Pilot

In 2012, through an open, Internet-based call for statements of interest, PCORI selected 33 volunteers to participate in a research prioritization pilot study. The participants included 16 researchers and 11 people who were patients, patient advocates, caregivers, or individuals from patient/caregiver advocacy organizations. The other six participants were stakeholders, such as clinicians, consumers, industry representatives, payer representatives, or policy makers. Dr. Rachael Fleurence, PCORI's Program Director for CER Methods and Infrastructure, stresses the importance of the patient perspective in the prioritization process. "If PCORI funds the study, the result of the research should allow patients to have information that matters to them and is actionable. By including patients and other stakeholders in the prioritization process, we probably will obtain a different set of topics."

The participants ranked 10 topics using a point system. They were asked to base this ranking on the following criteria: 1) patient-centeredness; 2) impact; 3) differences in benefits and harms; 4) reduction in uncertainty; 5) implementation in practice; 6) duration of information; 7) healthcare system performance; and 8) inclusiveness of different populations. "The pilot gave us a lot of information about how to improve our multi-stakeholder prioritization process," Fleurence says. "For example, eight is a lot of criteria, and pilot participants wanted to know if there was a way to streamline them." As a result, PCORI collapsed the prioritization criteria from eight to five: 1) patient-centeredness; 2) impact on population and individual health; 3) differences in benefits and harm, and reduction in uncertainty; 4) implementation in practice; and 5) duration of information.

On April 19 to 20, 2013, PCORI convened its first advisory panel meetings. Each of three stakeholder panels used the revised prioritization process to review between 10 and 25 topics to advise PCORI on key areas of research for the development of funding announcements. Fleurence concludes, "From the pilot, we saw that the process worked, and we knew that the process would work for the advisory panels."

Value-of-Information Analysis

VOI analysis may be used to identify questions that have the greatest potential to improve population health. VOI analysis is a process, rooted in statistical decision and economic theory, that projects the value of the findings of proposed research by estimating the average improvement in outcomes expected by obtaining the additional information. The process addresses the benefits and risks associated with treatment choices and the ability of research findings to reduce that uncertainty. VOI analysis takes into account the research-prioritization factors by integrating them into a single measure: the expected (average) increase in population health that might be expected from a research project (Meltzer et al. 2011; Meltzer 2001; Raiffa and Schlaifer 1961; Rein 2012; Meyers et al. 2012).

Peer and Stakeholder Review

Review of research proposals by scientists, patients, and other healthcare stakeholders is the final stage in selecting research proposals for PCORI funding. This review process identifies those proposals most likely to fulfill the institute's objectives and agenda (see **Patient Voices: PCORI Reviewers**). Despite its central role in scientific discourse and decision making, peer review of research proposals has had little attention as a subject of research. Rigorous experiments testing alternative approaches to peer review are rare; most peer-review practices are maintained by convention.

PCORI has particular advantages and responsibilities in developing its approach to peer review of research proposals. For example, incorporating patients and other stakeholders in peer-review activities presents a new opportunity but also makes the process more complex (Kotchen and Spelley 2012). Review practices vary substantially, and it is not possible to recommend one mode over another—or even to recommend when peer review of proposals is the best possible way to allocate funding and other resources. Nevertheless, independence between those being reviewed and those reviewing proposals must be safeguarded to maintain integrity of the process.



PATIENT VOICES: PCORI Reviewers

As part of “research done differently,” PCORI includes patients, caregivers, and other healthcare stakeholders in reviewing funding applications. PCORI has interviewed patient reviewers to learn more about this experience from their perspective, asking questions such as: Why did you apply to be a reviewer? What was most rewarding? What would you say to someone who has never been a reviewer before? and What would you say to patients who may feel intimidated about being a reviewer? Below are insights from two patient reviewers. These interviews and others are available at pcori.org/reviewerrecruitment.

“The whole purpose of doing patient-centered research is to benefit patients, and part of that is that we need participation from all people affected by healthcare … so part of that is going through technical documents and reviewing proposals and learning about research and science. But that’s accessible to anyone. I don’t think you need technical expertise, just need intelligence and integrity and the willingness to review the applications.” — *Caroline Leopold*

“[The] PCORI funding process was more streamlined. I was intimidated being side by side with scientific stakeholders, but I also felt like my input was valuable to the panel. Everyone on the panel wanted to hear my thoughts, and they appreciated what the patients were bringing to the panel because our experiences are so different than a scientist’s I found it to be a rewarding experience because I learned things from the other stakeholders, and I know that they learned things from me as a patient.” — *Crystal Brown Tatum*

Choosing a Study Design: Translation Framework

Even when the need for new PCOR has been clearly identified and prioritized, the quality of evidence provided by any new research is dependent on the design and quality of the studies conducted. The choice of study designs has practical implications for the timeliness, validity, and relevance of PCORI's research agenda. PCORI's [authorizing legislation](#)¹ directs the organization to develop "a translation table that is designed to provide guidance and act as a reference for the Board to determine research methods that are most likely to address each specific comparative clinical effectiveness research questions."

Very few published articles mention the concept of a "translation table," and researchers and stakeholders have varying opinions about what it should include (Tunis et al. 2012; Gliklich et al. 2012; Montori et al. 2012). Given this uncertainty, PCORI conceptualizes the translation table as a tool to guide the choice of study designs for specific research questions by balancing such factors as validity of the resulting evidence, appropriate use of scarce research resources, and timeliness of results. From this perspective, a translation table can balance the inherent tradeoffs of each study design and analytical methodology.

The challenge is that study design, as defined here, is multifaceted and complex. Although algorithms exist to help with aspects of study design—such as determining what statistical test should be used with what type of data, or specifying the data collection approach for different types of research—there is no formula that can be applied to all situations in PCOR. It is precisely for this reason that methodological expertise is often needed and why priorities, values, and available resources be considered when choosing a study design. After much discussion and input from several stakeholders, PCORI outlined a framework for the translation of research questions to study designs.

Translation Framework

The translation framework summarizes the important factors that underlie the construction of translation tables (see figure in [Appendix E: Translation Framework](#)). The translation framework begins with the patient's healthcare decision and assumes that a patient-centered research question has been precisely specified. Formulating decisions and defining questions for research are complex processes,

¹ Available at pcori.org/assets/PCORI-Authorizing-Legislation-032310.pdf

requiring expertise and open-mindedness (Bravata et al. 2005; Buedo 2005), careful formulation of research questions, and a commitment to patient-centeredness. The decision the study is meant to inform must be clearly defined, and a critical appraisal of prior studies must be undertaken. Multiple perspectives—including those of patients, clinicians, researchers, policy makers, and other stakeholders—may shape the research question. The components (often abbreviated PICOTS) of a well-formulated research question include:

- Population of patients/research participants and relevant subgroups of patients;
- Intervention(s) relevant to patients in the target population;
- Comparator(s) relevant to patients in the target population;
- Outcomes that are meaningful to patients in the target population, including the Timing of outcomes and length of follow-up; and
- Settings in which the intervention is delivered, including the healthcare providers.

Once the research question has been well defined, the design of appropriate studies needs to be considered. The translation framework is a process for making decisions about research designs that best provide a balance among such factors as the timeliness, resource requirements, and scientific rigor of alternative approaches. The framework includes the following guidelines for these important tradeoffs:

Keep the research question and the methodology separate: Any particular research methodology is the means to answer a research question as well as possible, but it is not a factor that should influence the choice of research question. Problems occur when the choice of a research question is driven primarily by data availability. Defining the question should not be limited by concerns about eventual methodological constraints. In PCOR, identifying decisions and defining a patient-centered research question should come first.

Focus on clarifying tradeoffs: After a research question is defined, choices have to be made about the type and level of evidence needed to inform the decisions it was intended to address. These choices will direct the research design and analytic strategy. The evaluation of alternative designs should be based on a series of factors, including timeliness, representativeness, and validity of findings, and the ability to identify subgroup effects. Such study characteristics (see **Examples of Study Characteristics**) substantially influence the usefulness of the results for decision making. Clearly articulating the tradeoffs among these characteristics will bolster the transparency in the selection of the analytical approach.

Place individual research studies in the context of a research program: Most research questions can be answered in several ways. A research program may, for example, include an effectiveness study based on existing healthcare data, a detailed interview study, and a randomized trial to balance population representativeness, timeliness, depth, and validity for informed decision making.

For example, suppose a new surgical procedure to repair a heart valve is less invasive than the standard surgery, but it requires specialized surgical training and skill and the participation of the cardiac surgery team. A randomized

trial may be required to establish the benefits or harms of the new procedure compared to the standard procedure under ideal conditions.

Regulators are likely to be very interested in the outcome of a study with this type of design. An observational study may also be needed to determine the safety and effectiveness of the new procedure, compared to the standard approach, when the procedure becomes more widely available.

Take into account advances in research methodology when choosing a study design: Over the past 20

years, choice of study design has been debated intensely in scientific and, more recently, political circles. These discussions often reiterate commonly held beliefs about randomized controlled trials (RCTs) and observational studies. Some people assert that RCTs are more relevant to decision makers than observational studies, and many RCTs have proven to have long-lasting value in clinical decision making. In many fields, critical evidence comes from RCTs, many of them conducted in patient populations and circumstances that are broadly applicable. Observational studies can have serious flaws that render them invalid and even irrelevant. However, well-designed observational studies have also been extremely valuable as a complement to RCTs, helping to determine under

Examples of Study Characteristics

Intrinsic Study Characteristics

- Internal validity (the extent to which effects are caused by the intervention or exposure)
- External validity (generalizability or applicability to non-study settings and populations)
- Precision (having small random error of estimation)
- Heterogeneity in risk or benefit (risks or benefits vary by subgroup)
- Ethical dimensions of the study (including considerations of risk–benefit balance and study burden for study participants)

Extrinsic Study Characteristics

- Timeliness (rapidly changing technology, policy, or public health needs)
- Logistical constraints (feasibility of collecting information from participants, number of participants available, study complexity)
- Data availability, quality, and completeness

what circumstances and to which patients the findings of RCTs apply. Serious errors in clinical practice can be due to overreliance on narrowly focused RCTs or on flawed observational studies, but both of these basic designs can contribute to answering clinical questions. Advances in research methodology can make RCTs more relevant, timely, and flexible, and they also can improve the validity of observational studies. In particular, the use of observational studies to make causal inference is potentially much stronger than it has been in the past. Some of the PCORI Methodology Standards address ways to improve the value of observational studies for questions about comparative clinical effectiveness. Decisions about study design need to take into account these standards and the advances in methodology they reflect.

When considering the tradeoffs among various methodological approaches, there are usually more than one acceptable design and analytic strategy. Choosing an appropriate study design involves tradeoffs among the limitations inherent in each design and analysis approach. For example:

- To obtain results sooner or to maximize external validity, an observational study using secondary data (information from previously collected data) could be considered. However, this design would likely have less internal validity than would an experimental study that uses randomization. The experimental study could fail to address the research question, though, if it is not representative of care outside the controlled research environment. In contrast, a study design without a comparator, based on information from a device registry, *might* be acceptable for assessing device failure rates but not for assessing device effectiveness.
- Often logistical issues can be more challenging than scientific ones. For example, if only a limited number of patients with a specific condition are available to study, then an efficient sampling strategy within existing healthcare data sources could facilitate study success.

Once a decision is made to conduct an observational or experimental study, a number of options about study design need to be considered and weighed. Some treatments are used sporadically or just once. If the treatment of interest is short term or one time (e.g., antibiotics, vaccines), then self-controlled designs offer attractive properties, including improved confounding adjustment for time-invariant patient factors. In approach and interpretation of results, self-controlled designs are sufficiently different from other nonrandomized studies such that additional issues need to be considered. If an exposure of interest is longer term, then a cohort study design needs to be considered.

Variation in exposure to the treatment also needs to be addressed. This issue includes considerations of whether variations in exposure stem from other, higher-level causes. For example, providers may have a

strong preference for one treatment over another. Such variation can be observed on a regional level, depending on insurance constraints, or over time. Sometimes preference for a treatment changes rapidly after new medical evidence arises. Such variation can be exploited using time-trend analyses or instrumental variable analyses, both of which may provide advantages in confounding control. Finally, choices must be made regarding the most appropriate data sources. Does the nature of the study question require that specific information be newly collected, or will information from previously collected data suffice? Within the domain of previously collected data, several factors need to be considered, including clinical detail, data completeness, access to the data, and confidentiality. Often the linkage of multiple data sources is most promising.

The Translation Table

The translation framework described above provides the foundation for a method to summarize the trade-offs in choosing which research designs could provide valid and useful information to fill clinical evidence gaps in a timely fashion. The translation table, which is the means for providing this summary in a standardized way, is created individually for each important clinical question being considered for new research. The translation table fosters discussion among research planners about the tradeoffs and includes ratings for each important factor as applied to each alternative design. Thus, the table is a communication tool to guide choices rather than an algorithm for specifying the preferred choice. Although the translation table can be designed in different ways, its purpose is to summarize the strengths and weaknesses of alternative study designs for any particular research question. The table lists various criteria (i.e., internal validity, generalizability, logistics, and the required resources) and provides ratings for all criteria for each study design that would be considered. From this perspective, the translation table is a template for summarizing the criteria included in the translation framework as they are applied to specific clinical questions.

SECTION III: PCORI METHODOLOGY STANDARDS

Introduction

Because patient-centered research outcomes (PCOR) can include a variety of research designs and specific techniques, PCORI's initial set of 47 methodology standards are broad and do not address all possible issues in clinical research. The topics for the standards were chosen to reflect areas where: 1) there were either substantial deficiencies or inconsistencies in how available methods were applied in practice, despite specialized knowledge in how best to conduct research; or 2) there were barriers to the effective dissemination of research results (Helfand et al. 2011; Lohr 2007; Schneeweiss et al. 2012). In July 2012, the standards were released in draft form and public comments were solicited. The standards were then revised in response to the public comments. PCORI's Board of Governors endorsed the revised standards, which were released to the public in December 2012. Details on the standards development process are provided in **Appendix D: The PCORI Methodology Committee's Approach to the Authorizing Legislation**. The background papers commissioned to guide development of the standards, the draft report, and public comments about the draft report and standards are available on [PCORI's website](#).¹

Over the past four decades, explicit, formal standards for planning, conducting, and reporting clinical trials were developed for the subset of research studies that are conducted to obtain regulatory approval from the US Food and Drug Administration (FDA 2010 a, b). These standards, articulated in formal "guidance documents," helped to create a level playing field for companies designing such studies and for regulatory decision makers. PCORI's Methodology Standards are not intended to replace the FDA guidance documents, nor has PCORI requested that FDA adopt its standards. Rather, these new standards are meant to provide guidance to the broad community of researchers who conduct PCOR.

The PCORI Methodology Standards specifically address the design and conduct of PCOR studies, distinguishing them from ongoing efforts in the past decades to develop reporting standards for studies that employ certain designs. These guidelines are currently housed at the [Equator network](#)

¹ Available at www.pcori.org/research-we-support/research-methodology-standards

website², which includes widely utilized tools such as CONSORT (for randomized clinical trials), STROBE (for observational studies), and STARD (for diagnostic accuracy studies).

In 2008, the Institute of Medicine (IOM) stated that methodological standards for the conduct of one type of research—systematic reviews—would help decision makers “with respect to transparency, minimizing bias and conflict of interest, and clarity of reporting” (IOM 2008). In 2011, the IOM published standards for conducting systematic reviews (IOM 2011). The PCORI Methodology Standards extend this work by formulating standards for comparative clinical effectiveness research including randomized trials, observational studies of effectiveness, and studies of diagnostic tests.

As a group, the PCORI Methodology Standards offer an approach to aligning a PCOR program with research questions that can address patients’ and clinicians’ uncertainty about what works best, for whom, and under what circumstances. Methodological standards can improve the way research questions are selected and formulated, how studies are designed to address these questions, and how findings are reported. Standards can also help prevent the use of flawed methods. Just as standards helped to define the quality of evidence required for decisions about regulatory approval of a new drug or device, standards for PCOR can benefit medical innovators by providing a common set of expectations about the characteristics of high-quality research.

The PCORI Methodology Standards consist of 47 individual standards (see **Appendix A: PCORI Methodology Standards**). This initial set of standards is necessarily incomplete, addressing an initial group of topics chosen to contribute to the quality and value of PCOR. These standards represent the first phase in PCORI’s ongoing effort to promote research methodology that will support the design and conduct of high-quality PCOR.

The 47 individual standards fall into 11 categories. This report discusses each of those categories and summarizes the justification for the related standards.

² Available at equator-network.org

The first five categories are cross-cutting and are relevant to most PCOR studies. Researchers should refer to all of these standards when planning and conducting their research projects. These categories are:

- Formulating research questions
- Patient-centeredness
- Data integrity and rigorous analyses
- Preventing and handling missing data
- Heterogeneity of treatment effect (HTE)

The other six categories of standards are applicable to particular study designs and methods. Two of the categories provide guidance on developing specific types of data and using them in studies:

- Data registries
- Data networks as research-facilitating infrastructures

The final four categories of standards apply to studies that have varying study designs and purposes. The standards in these categories should be used for guidance when relevant to a particular study:

- Causal inference methods
- Adaptive and Bayesian trial designs
- Studies of diagnostic tests
- Systematic reviews

Most of the standards should be considered minimal standards, meaning that they are necessary for sound science but should not inhibit further evolution of methods. Some standards are designed to promote transparency: how to communicate properly, both in study protocols and in published reports, exactly what was planned and what was done. All the standards are based on current scientific knowledge, either from empirical studies when they were available or from theoretical work about research methods.

In the following sections, the standards are grouped by category. The sections include a brief summary of the rationale for the standards and the full text of all standards, with key definitions. References to the applicable standard are included in parentheses, e.g., (RC-1), and a list of the standards appears at the end of each subsection.

Cross-Cutting Standards for PCOR

1. Standards for Formulating Research Questions

Research involves four broad phases or categories of activities:

- “What should we study?”
- “What study designs should we use?”
- “How do we carry out and govern the study?”
- “How do we enable people to apply the study results?”

Many of the PCORI Methodology Standards focus on the early phases of research because getting the questions right (“What should we study?”) is an important starting point. The standards specify what to include in research protocols as a means of increasing study quality as well as transparency in research. The intended results are both better studies and a better understanding of the applicability of study results to specific patients and situations.

Rationale for the Standards

All high-quality, useful research begins with good planning. For PCOR, these planning steps are necessary to ensure that the research will be relevant to clinical decisions, that recruitment strategies will achieve participant numbers required for scientific rigor, and that the protocol makes clear how the research will accomplish its objectives.

To ensure the PCOR is relevant to decision making, the need for a new study must be rigorously justified. Investigators should identify the gaps in evidence their study will address (**RQ-1**). Given that resources are limited, study questions should not be redundant or irrelevant to practice and decisions. One way to avoid these pitfalls is to identify an existing systematic review or conduct such a review before pursuing additional research (Ransohoff 2007).

Once the need for new research is established, a formal study protocol should be created (**RQ-2**). In addition to defining the clinical decision being addressed, research protocols present comprehensive plans for how the research will be executed. Formal protocols make the study intentions clear to all

users and provide the details that are needed to evaluate quality of the research. In addition, they ensure that spurious results are not reached as a result of multiple post hoc analyses.

Identifying who will be in the study population is essential to understanding to what patients the results will apply (**RQ-3**). Many studies also aim to define how the treatments being compared impact significant subgroups of the population. These subgroups should be specified, along with enough detail about the sample size and participants to evaluate any subgroup differences reported in the results.

Standards **RQ-4** through **RQ-6** provide the minimal requirements of a detailed protocol, which include specifying the subgroups, interventions, comparators, and outcomes. A more in-depth discussion of the selection of patient-centered outcomes can be found in the section on Standards Associated with Patient-Centeredness.

Cross-Cutting Standards for PCOR

1: Standards for Formulating Research Questions

RQ-1 Identify gaps in evidence

Gap analysis and systematic reviews should be used to support the need for a proposed study. If a systematic review is not available, a systematic review should be performed using accepted standards in the field (see standard SR-1), or a strong rationale should be presented for proceeding without a systematic review. In the case where a systematic review is not possible, the methods used to review the literature should be explained and justified.

RQ-2 Develop a formal study protocol

Studies should include a formal protocol specifying at least one purpose for which the data were collected (e.g., effectiveness, safety, natural history of disease, quality improvement); data sources and linkage plans, if any; data feasibility and quality, measure(s) of effect; and use of any standardized data dictionaries (nationally or internationally accepted).

RQ-3 Identify specific populations and health decision(s) affected by the research

To produce information that is meaningful and useful to people when making specific health decisions, research proposals and protocols should describe: 1) the specific health decision the research is intended to inform; 2) the specific population for whom the health decision is pertinent; and 3) how study results will inform the health decision.

RQ-4 Identify and assess participant subgroups

In designing studies, researchers should identify participant subgroups of interest and, where feasible, design the study with adequate precision and power to reach conclusions specific to these subgroups. In addition, subgroup information should be reported for later systematic reviews.

RQ-5 Select appropriate interventions and comparators

When evaluating an intervention, the comparator treatment(s) must be chosen to enable accurate evaluation of effectiveness or safety compared to other viable options for similar patients. Researchers should make explicit what the comparators are and how they were selected, focusing on clearly describing how the chosen comparator(s) define the causal question, reduce the potential for biases, and allow direct comparisons. Generally, non-use (or no specific treatment) comparator groups should be avoided unless no specific treatment is a likely option in standard care.

RQ-6 Measure outcomes that people representing the population of interest notice and care about

Identify and include outcomes the population of interest notices and cares about (e.g., survival, function, symptoms, health-related quality of life) and that inform an identified health decision. Define outcomes clearly, especially for complex conditions or outcomes that may not have established clinical criteria. Provide information that supports the selection of outcomes as meeting the criteria of “patient-centered,” and “relevant to decision makers” such as patient and decision maker input from meetings, surveys, or published studies. Select outcomes based on input directly elicited from patient informants, people representative of the population of interest, either in previous studies or in the proposed research.

2. Standards Associated with Patient-Centeredness

The purpose of PCOR is to help people make informed healthcare decisions. To do this, PCORI must direct research toward asking questions that are important to patients, measure outcomes that are noticeable and meaningful to them, and produce results that help them weigh the value of healthcare options given their personal circumstances, conditions, and preferences.

To conduct PCOR, researchers must engage people representing the population of interest and other relevant stakeholders in the design, conduct, and dissemination phases of research. For some populations—for example, children or cognitively impaired persons—representatives may include surrogates and caregivers. To inform patient decisions, PCOR must accurately incorporate patient needs, values, and preferences into the procedures chosen for conducting all parts of the study. While all the standards are designed to advance high-quality PCOR, the standards in this group directly promote effective patient engagement and the explicit incorporation of patient needs, values, and preferences into research. Engagement of patients can meaningfully contribute to several stages in a research project including:

- Defining topics and formulating study questions;
- Identifying a study population and choosing interventions, comparators, and outcomes;
- Developing optimal strategies for recruitment and retention of study participants;
- Conducting a study and analyzing results; and
- Disseminating research findings into clinical practice.

Because currently there is only limited evidence about the impact of patient involvement on the quality of research, these standards are based on consensus about the best ways to promote research that is patient-centered. While there are some guidelines and recommendations regarding patient engagement in research (Deverka et al. 2012; Staniszewska et al. 2011), the empirical evidence underlying the standards varies considerably in quality and quantity (Staniszewska et al. 2011; Gagnon et al. 2011). Nevertheless, PCORI's standards associated with patient-centeredness are designed to identify optimal approaches for engaging patients and other stakeholders throughout the research process and to gain a better understanding of how such engagement affects study design and outcomes.

Without prescribing a specific approach, the standard **PC-1** directs researchers to formulate and describe their methods of patient engagement. This engagement can include involving patients in developing questions and defining participants, interventions, and comparators, as appropriate. Patient engagement comprises activities that are fundamentally different from the conventional concept of enrolling patients as participants in clinical research studies (see **Patient Voices: Lucinda Shore** and **Research in Practice: Pamela Williams**).

PCOR also requires that study participants be representative of the spectrum of the population facing the health decision of interest. For this reason, the standards require that research proposals and reports document how the researchers identify, recruit, and retain study participants (**PC-2**). In developing standard **PC-2**, PCORI evaluated specific strategies for involving people who have been historically underrepresented in research or who are considered to be hard to reach (Mullins et al. 2012).

An explicit focus on patient-centered outcomes is a defining characteristic of PCOR (**PC-3**). As suggested by Guyatt and colleagues, an outcome that is relevant to patients must pass the following test: “Were it to be the only thing that changed, patients would be willing to undergo a treatment with associated risk, cost, or inconvenience” (Guyatt et al. 2008). Many (though not all) meaningful and important patient-centered outcomes, such as symptoms, are best reported by patients themselves, and these are called patient-reported outcomes (PROs). Pain and some other outcomes cannot reliably or accurately be assessed by any means other than direct patient report, so inclusion of PROs is often essential to patient-centeredness.

The standards require the use of validated PRO instruments when they are available. At the same time, the standards encourage development and testing of new PROs, if needed, to measure the outcomes that are important to patients (see **Patient Voices: Juli** and **Patient Voices: A Woman with Fibromyalgia**). It is crucial to work with engaged patients to identify new measures that reflect what is significant to them. In some circumstances, PRO instruments can be adapted to each individual, based on their preferences and goals for treatments).

To complete the research continuum from the patient-centeredness perspective, dissemination of the study’s findings should integrate the new results with related work and underscore meaningful clinical

and policy implications. While dissemination may be outside the scope of an individual research project, researchers should support the ultimate dissemination and implementation of their results. They can do this in several ways, including presenting results in formats that are accessible and understandable to target audiences, such as clinicians, patients, and caregivers (**PC-4**).



PATIENT VOICES: Lucinda Shore

Nine years ago, Lucinda Shore noted episodes of shortness of breath and chest pain punctuated by rapid breathing and anxiety. She reported this to her doctor, and for the next five years, was misdiagnosed with conditions ranging from stress to hormone imbalance to heart disease. Shore finally learned that she had emphysema from a genetic disorder called Alpha-1 Antitrypsin Deficiency, often called simply Alpha-1. Today, at age 49, Shore receives weekly infusions of an enzyme she is missing; the treatment slows the progression of the disease and keeps her damaged lungs from deteriorating further. She expects to require such augmentation therapy for the rest of her life.

Shore is a patient partner in the PCORI Pilot Project whose goal is to document the social and psychological health outcomes that affect people with rare diseases—illnesses found in fewer than 200,000 patients in the United States. The project aims to develop a measurement tool that defines the way these diseases affect a patient's life beyond the medical symptoms. Shore's experience with her delayed Alpha-1 diagnosis and treatment and her desire to push physicians to see “the big picture”—and thus provide better care for patients—is a major incentive for her participation in the research project. The many psychosocial issues and day-to-day challenges associated with a chronic disease are of particular concern to Shore. These include the stigma of having a chronic condition, the fear that her sons will also develop it, a mistrust of doctors after her years of receiving incorrect diagnoses, and difficulty in social situations, such as dating. “When do you tell a person that you have a genetic disease?” Shore asks. “If I become extremely short of breath, it is concerning for people to hear me breathe. They wonder if I’m dying,” Shore says.

Among her project activities, Shore has helped seek out other patient partners and recruit participants. She also conducted a focus group with patients. She currently works on data analysis and is in regular contact with researchers about the project’s progress. Shore believes including patient partners in a research project can offer researchers a different and valuable perspective. She says of her experience leading a patient focus group: “Patients speak with doctors and clinicians about certain issues, but when you’re around someone else who has your same condition, you tend to open up and you tend to share issues with each other that you don’t necessarily share with your doctor.”



RESEARCH IN PRACTICE: Pamela Williams

Millions of Americans with rare diseases often deal not only with misdiagnoses, diagnostic delays, and a frustrating search for treatments, but they may experience social and psychological problems the health care system doesn’t recognize. Pamela Holtzclaw Williams, PhD, JD, RN, wants to change that. Williams, University of Arkansas researcher, was awarded a PCORI contract to use feedback from patients with the rare disease Antitrypsin Deficiency (Alpha-1) to tailor instruments to develop social burden measurement tools that are adapted by and for the Alpha-1 community and others with rare diseases. Alpha-1 is a genetic disease that causes serious liver disease in children and liver and lung disease in adults.

“We’re trying to measure the social determinants of health,” Williams says, assessing things like access to competent care, access to medicines, length of time to diagnosis, burdens of the disease, and a series of decisional burdens. Williams has formed a community-based participatory research partnership with the Alpha-1 community, which has a vibrant nationwide patient advocacy network in place. “People [with Alpha 1] are telling us new categories that can be included in [our] instruments,” Williams says. Decisional burdens faced by those with rare genetic illnesses include factors such as: Who gets tested in the family? Who should learn receive the results? Should they get married? Should they have children?”

Community partners, who sit on an advisory board that meets monthly, have been instrumental in recruitment of not just partnership members but also study participants from the community. Being a patient and community partner is not just a token leadership role. “My patient and community partners have told me that participating in the research project has made them have a better focus in their advocacy work; they are learning how to be strategic about their expenditure of energy,” Williams says.

While there have been challenges to her research—specifically, finding training for community partners on the particular processes common to a research environment, such as the technicalities of institutional review boards and grant writing, Williams has found the collaboration with patient participants overwhelmingly positive. Williams believes that patients should be a part of the research process from start to finish and that other researchers need to know that while it takes time and patience to collaborate with patient and community partners in research, the outcomes are beneficial to both the patient and research communities. “It’s important to keep the project relevant to the patient-centered outcomes,” Williams says, “as opposed to being focused and relevant to institutional or providers’ desired outcomes.”



PATIENT VOICES: Juli

When Juli was diagnosed with breast cancer, she worked through her options with her primary care doctor, Leigh Simmons, MD. Juli had extensive cancer in her left breast that had spread to her lymph nodes and her right breast. With her doctor, Juli made the decision to proceed with a double mastectomy.

Juli says, “My decision, perhaps as for most breast cancer women, was very simple. I have breast cancer in both; if one is coming off, the other is coming off.”

Having decided to proceed with the mastectomy, Juli and Simmons put together a treatment team comprised of an oncologist, a surgical oncologist, a plastic surgeon, a radiation oncologist, nurse practitioners, and nursing staff. “You realize these people are going to be very important for the rest of your life,” Juli says. “They’re going to be explaining things that I didn’t have a whole lot of knowledge about. I’m going to have to do a lot of research. I’m going to have to depend on them.”

Even though Juli had decided on a course of action, she still had questions and reservations about her treatment and expected outcomes and looked to Simmons to help communicate this. One outcome that was of particular importance to Juli was her ability to continue to play bagpipes.

“Not only was it, ‘Oh, I want to play my music,’ but it’s a great distracter for me,” Juli says. “It’s a great comfort for me to get out with my band and to play.”

Simmons says, “I really hadn’t thought about how that was going to be a problem after surgery, but she explained to me that there was potential that it might be because of where she holds the pipe.” She was reminded that the point of being treated for cancer is to enable the patient to continue to live a full life.

When she and Juli met with the treatment team, they were able to communicate the importance of this outcome for Juli’s health and wellbeing. The team listened and worked to set up a course of action that would have the least possible impact on her ability to play bagpipes.

“It didn’t eliminate [the issue], it still had some impact,” Simmons says. “But they really heard what she was trying to say and they realized that unless they kept [in mind] her needs to be able to do the things that she needed and loved to do, if they didn’t get that part right, the rest of her treatment might not go as well either.”



PATIENT VOICES: A Woman with Fibromyalgia

Fibromyalgia is a condition characterized by widespread pain.

An MRI cannot tell a physician how my pain affects me. An EMG cannot tell a physician how severe my pain is. A blood test cannot tell my physician what challenges I face.

On my first and subsequent visits to my rheumatologist, I was asked to fill out a questionnaire about my feelings and thoughts about my pain. My rheumatologists' office used a questionnaire called the Multi-Dimensional Health Assessment Questionnaire (MDHAQ). The questionnaire asks 13 questions about what you have been able to do over the past week and uses the scale "without any difficulty," "with some difficulty," "with much difficulty," and "unable to do." It asks questions such as: Am I able to dress myself? Get in and out of bed? Lift a full cup or glass to my mouth? Bend down to pick up clothing from the floor? Walk two miles? Participate in sports and games as I would like. With the exception of participating in sports and games as I would like, I am capable of doing everything on this questionnaire without any difficulty.

The activities listed on the questionnaire do not encapsulate my life, and they do not include activities that are difficult for me. I have difficulty picking up heavy or oddly shaped items. I have difficulty opening bottles. I have difficulty dancing. I have difficulty sitting for long periods of time. I have difficulty lying down. I have difficulty holding my 20-pound niece when she's asleep in my arms. How can this questionnaire monitor my physical limitations and improvements if it doesn't include activities or tasks with which I would have difficulty?

The MDHAQ also asks, on a scale of 0 to 10, how much pain I have had because of my condition over the past week. I was also asked to rate my pain on a 0 to 10 scale by orthopedic surgeons and physical therapists. When I first started rating my pain, my ratings were somewhat arbitrary. Rarely, if ever, did I say my pain was above a 3. This was not because my pain wasn't bad or didn't affect me, but because I wanted to be strong and not give in to the pain. I thought "I'm a strong woman with a high pain threshold. The pain isn't that bad."

It wasn't until I had a conversation with my cognitive behavioral therapist that we realized that my thinking about my pain was a little off for two reasons. First, I consistently underrated my pain. I did not truly understand how to distinguish a 2 from a 5 on the pain scale. How can I rate my pain a 2 if I need to stop what I am doing to address the pain? How can I call my pain a 2 if it interferes with my life and day-to-day tasks and if my focus shifts from the task at hand to my pain? Second, there was no consistency to my ratings, and my responses were a moving target from week to week, and not because the pain was different from week to week. My responses were not truly anchored or grounded in any symptomatology or experiences to allow for consistency.

Cross-Cutting Standards for PCOR

2: Standards Associated with Patient-Centeredness

PC-1 Engage people representing the population of interest and other relevant stakeholders in ways that are appropriate and necessary in a given research context

People representing the population of interest include individuals who have the condition or who are at risk of the condition and, as relevant, their surrogates or caregivers. Other relevant stakeholders may include clinicians, administrators, policy makers, or others involved in healthcare decision making. Stakeholders can be engaged in the processes of:

- Formulating research questions;
- Defining essential characteristics of study participants, comparators, and outcomes;
- Identifying and selecting outcomes that the population of interest notices and cares about (e.g., survival, function, symptoms, health-related quality of life) and that inform decision making relevant to the research topic;
- Monitoring study conduct and progress; and
- Designing/suggesting plans for dissemination and implementation activities.

When applicable, research proposals should describe how these stakeholders will be identified, recruited, and retained. If engagement is not necessary or appropriate in these processes, explain why.

PC-2 Identify, select, recruit, and retain study participants representative of the spectrum of the population of interest and ensure that data are collected thoroughly and systematically from all study participants

Research proposals and subsequent study reports should describe: 1) the plan to ensure representativeness of participants; 2) how participants are identified, selected, recruited, enrolled, and retained in the study to reduce or address the potential impact of selection bias; 3) efforts employed to maximize adherence to agreed-on enrollment practices; and 4) methods used to ensure unbiased and systematic data collection from all participants.

If the population of interest includes people who are more difficult to identify, recruit, and/or retain than other study populations (for example, individuals historically underrepresented in healthcare research such as those with multiple disease conditions, low literacy, low socioeconomic status, or poor healthcare access, as well as racial and ethnic minority groups and people living in rural areas), then specify plans to address population-unique issues for participant identification, recruitment, and retention.

PC-3 Use patient-reported outcomes when patients or people at risk of a condition are the best source of information

When patients or people at risk of a condition are the best source of information regarding outcomes of interest, then the study should employ patient-reported outcome (PRO) measures in lieu of, or in addition to, measures derived from other sources. Proposals should describe: 1) the concept(s) underlying each PRO measure (e.g., symptom or impairment) and how it is meaningful to, and noticed by, patients in the population of interest; 2) how the concept relates to the health decisions the study is designed to inform; 3) how the PRO measure was developed, including how patients were involved in the development; and 4) evidence of measurement properties including content validity,

2: Standards Associated with Patient-Centeredness (Continued)

construct validity, reliability, responsiveness to change over time, and score interpretability, including meaningfulness of score changes in the population of interest with consideration of important subgroups. If these measurement properties are not known, a plan for establishing the properties must be provided. Caregiver reports may be appropriate if the patient cannot self-report the outcomes of interest. If PROs are not planned for use in the study, justification must be provided.

PC-4 Support dissemination and implementation of study results

Support dissemination and implementation of study results by suggesting strategies, indicating clinical and policy implications, and working with patients or organizations to report results in a manner understandable to each target audience.

3. Standards for Data Integrity and Rigorous Analyses

The standards that address data integrity and rigorous analyses build on best practices in clinical research and add to the Standards for Formulating Research Questions by requiring documentation of key decisions and tests of the assumptions made in the analyses. These standards are applicable once a researcher has decided to use a specific research design or analysis methodology, and they apply to most study designs.

Rationale for the Standards

Data to be used for PCOR need to contain all the variables required by the proposed analyses. This is particularly important in observational studies that use preexisting data. Assessing data adequacy involves determining whether the data include the necessary information about other factors that could affect results (e.g., mitigating and confounding factors). It is also imperative to determine whether data on the important outcomes are available and valid (**IR-1** and **IR-2**).

Users of the research findings need to be able to evaluate whether the study produced accurate results and whether the results apply to their situations. Therefore, researchers must describe how the analyses were designed and conducted (e.g., data collection activities, settings, analytic techniques, means of assuring data quality, comparability of study groups) (**IR-3**). Such standards are

essential for transparency and scientific rigor, as they allow stakeholders to evaluate both the quality of studies and their applicability.

When data are derived from tests or scales, the test or scale characteristics as well as evaluations of their performance (psychometric properties) should be established and reported (**IR-4**). This provides a clear understanding of what researchers intended to measure and allows comparisons to be made across studies.

All research requires assumptions during data analyses, and these assumptions determine whether inferences are valid. Incorrect assumptions have the potential to invalidate a study's results. For this reason, assumptions need to be tested to the extent possible, not simply stated. Certain kinds of assumptions—particularly those that are central to an inference and cannot be tested directly using the study data—should be subjected to sensitivity analyses (**IR-5**). Sensitivity analyses repeat the analyses under different structural assumptions and then compare the results to see if the conclusions change materially.

In addition to the requirements of previously mentioned standards, researchers should provide sufficient information to permit assessment of the likelihood that their research results are true (internal validity) and would be the same in another group of participants (external validity) (**IR-6**). Guidelines for reporting on studies with various designs have been established by medical journal editors and other professional groups, and researchers who conduct PCOR should follow such guidelines, so that their work can be assessed and compared to other studies.

Cross-Cutting Standards for PCOR

3: Standards for Data Integrity and Rigorous Analyses

IR-1 Assess data source adequacy

In selecting variables for confounding adjustment, researchers should assess the suitability of the data source in terms of its ability to assure robust capture of needed covariates.

IR-2 Describe data linkage plans, if applicable

For studies involving linkage of patient data from two or more sources (including registries, data networks, and others), describe 1) each data source and its appropriateness, value, and limitations for addressing specific research aims; 2) any additional requirements that may influence successful linkage, such as information needed to match patients, selection of data elements, and definitions used; and 3) the procedures and algorithm(s) employed in matching patients, including the success, limitations, and any validation of the matching algorithm.

IR-3 A priori, specify plans for data analysis that correspond to major aims

Researchers should describe the analytic approaches that will be used to address the major research aims prior to data collection. These include definitions of key exposures, endpoints, and covariates. Also identify patient subgroups of interest , plans (if any) for how new subgroups of interest will be identified or how analysis plans may be adapted based on changing needs and scientific advances, and plans for how missing data will be handled.

IR-4 Document validated scales and tests

Studies should include documentation of the name of the scales and tests selected, reference(s), characteristics of the scale, and psychometric properties.

IR-5 Use sensitivity analyses to determine the impact of key assumptions

The results of these sensitivity analyses should be reflected in the interpretation of results.

IR-6 Provide sufficient information in reports to allow for assessments of the study's internal and external validity

Reporting guidelines for specific designs can be found at the EQUATOR Network website (www.equator-network.org). This website has brought together all reporting guidelines that have been developed using formal approaches, many of which have been adopted by journals, such as CONSORT (for randomized clinical trials), STARD (for diagnostic tests), and STROBE (for observational studies).

4. Standards for Preventing and Handling Missing Data

Missing data are unrecorded data values or unavailable information that would be meaningful for dataset analysis. A central aspect of planning clinical research is defining the set of data that will best meet the project’s objectives; however, final datasets are usually incomplete. One reason for missing data is errors in measurement or recording. Additionally, datasets derived from records not intended for research—such as those generated from routine clinical care—are particularly prone to missing data. Similarly, studies that involve patient populations that are harder to retain over time are likely to be missing data. Finally, missing data are virtually inevitable in studies of people. Missing data can be at the respondent level (“unit nonresponse,” in which a respondent chooses not to provide data) or at the variable level (“item nonresponse,” in which a respondent chooses not to answer a specific question). Both types of nonresponse are problematic, though unit nonresponse generally has more impact. Data may not be recorded because of patient actions unrelated to the study itself, such as missing a scheduled follow-up or dropping out of the study altogether. If proper statistical methods for handling missing data are not employed, the analyses of those data can be biased or can overstate the precision of the findings. The standards for preventing and handling missing data apply to these types of situations. The current standards do not cover cases, called “missing by design,” where data is not available because the study design did not include plans to record them.

Missing data is a particularly important consideration for PCOR, because such research often includes diverse patients and is conducted in diverse clinical settings. This variety can make collecting complete sets of data more challenging. Patients with more than one disease condition and those seen in community care settings may be more likely to be lost to follow-up than other patients. The effort to prevent missing data is one of several reasons researchers may choose to conduct studies in specialized clinical settings and exclude patients who, because of other clinical problems, might be less likely to complete the study. Consequently, the research may fail to represent actual results that would occur in more varied clinical settings and among more diverse patient groups.

Rationale for the Standards

Preventing and planning for missing data—and describing the methods used to address missing data in a study protocol—are minimal requirements for good research. Tracking all study patients and recording not only that a participant dropped out but also the reasons for dropout and loss to follow-

up is currently considered good practice and is required by many of the organizations that fund research and the journals that report results. The extent and pattern of missing data must be reported so that the implications are clear to anyone who might base a decision on the results.

Different patients face different challenges in participating in research studies (see **Patient Voices: Sarah**). Involving patients during the design of study can help identify and address potential reasons people might drop out of a study or data might be hard to collect. Researchers and patients should work together to identify and address those reasons (**MD-1**). Informing patients about the implications of missing data may help researchers retain study participants and meet the intent of the Standard **MD-1** (see **Research in Practice: Missing Data**).

Many researchers and groups have provided guidance on the handling of missing data (National Research Council 2010). The science and analytic software now available can facilitate several different rigorous approaches to handling missing data (**MD-2**).

To reduce the risk of selecting an approach that could adversely affect either the validity or the relevance of the study results, researchers—before seeing the data—should determine how the analysis will address missing data. In the past 30 years, many new methods for handling missing data have been developed. Some may require statistical expertise. Methods that use multiple values for the missing value are more likely to produce accurate results and should be used in most situations rather than using a single value (e.g., the baseline or the last observation carried forward) or only including cases with complete data in the analyses (see **Research in Practice: Bias in the Last Observation Carried Forward Method; MD-3**).

All missing data methods rely on assumptions that are related to the study topic and design. Three common assumptions about the impact of missing data are:

- What is missing has nothing to do with a patient’s characteristics (known as “missing completely at random”);
- What is missing depends on patient characteristics predictive of the outcome, and these characteristics were measured (“missing at random”); or
- What is missing depends on patient characteristics predictive of the outcome that were either not measured or not observed (“missing not at random,” or “non-ignorable” missingness).

To evaluate and select the appropriate assumption, it is important to have as much information about missing data and the patients lost to follow-up (**MD-4**)., The assumptions about the causes of missing data are important in determining what methods to use for the analyses. When researchers make one of these assumptions and then base their choice of analysis methods on it, they should consider how making a different assumption would affect their results (referred to as assessing the sensitivity of inferences) (**MD-5**). This is especially important if the amount of missing data seems likely to affect the study results.



RESEARCH IN PRACTICE: Missing Data

Courtney Schreiber, MD, MPH, is a gynecologist and clinical researcher at the University of Pennsylvania School of Medicine. Here she discusses how she uses patient narratives to learn more about how to tailor her studies to the needs of patients. She also uses her patient stories to help recruit and retain enrollees in clinical trials.

How do you talk about missing data with patients?

Schreiber: I often tell a story about a participant named Sally. She enrolled in one of our contraceptive clinical trials. She was absolutely committed to helping women like herself figure out which type of contraception is best. But, after a while, she stopped coming to her study appointments for a logistical reason. When we called her up, she had no idea that dropping out of the study would make it harder for us to learn which medicine worked best. She knew that other women were waiting to enroll in the study, so she thought that someone could just take her spot.

Did Sally leave the study?

Schreiber: No. We were able to figure out how to get her to her appointments: by keeping the research office open late on Thursday. One of the key factors in keeping Sally was being able to show her how much harder it was for us to figure out which medication worked best if we didn't know how she felt at the end of the study. She had been feeling pretty good and thought we could just use the data we had. But once Sally was able to understand how helpful it was for her to stay on as part of the team, she finished the whole study.

How is Sally's story useful in retaining participants on other studies?

Schreiber: We always promise our study participants that we will work with them to find the most convenient ways to participate, but that message doesn't always stick. But many of them identify with Sally's story, so it helps us explain why staying in the study is so helpful. And it really seems to work.



PATIENT VOICES: Sarah

Sarah is a 61-year-old retired hospital clerk living in the UK. She is married and a mother of two grown children. In 2002, after seeing a recruitment flier posted in the hospital where she worked, Sarah volunteered for a placebo-controlled clinical trial intended to help women at risk of osteoporosis.

Because she had broken several bones in the past and was over 50 years old, Sarah felt she might be at risk for osteoporosis. A body scan confirmed that Sarah did have osteoporosis, and so she began the trial regimen which involved injecting the trial drug, or a placebo, into her abdomen twice daily. Besides being interested in the benefits she might individually receive from the trial, Sarah felt it was important to join the trial to help others.

“All you can say is you’re doing your best to help other people and mankind, and we won’t get anywhere if nobody volunteers for anything,” Sarah says. “And it may give you some benefits. At least you know in your mind, you’ve done something to help people. And if there aren’t that many of you with the illness, et cetera, it’s very important you volunteer.”

As Sarah began the trial, she found the injections were very difficult to handle. She found the injections to be a painful and a nuisance, which she came to dread. “Every day, I had to steel myself to do it. I’ve got a bit of a big tummy anyway, but I could still feel everything: taking a lump of stomach, swab it, of course, and—oh, I don’t know—it’s making my mouth go dry. I don’t know if it’s fear or what, but I was doing that, for months before I realized that I really, really could not cope any longer.”

Yet, Sarah continued with the trial despite her discomfort. “I get myself so far into things; I don’t like to back out. I didn’t want to disappoint [the nurse] because she was saying ‘Oh, it’s wonderful you’ve come forward, so few people have.’” However, after visiting a very ill relative in the hospital, Sarah found she related the smell of the hospital with her experience in the osteoporosis drug trial. She realized she could no longer cope with the study and decided to withdraw.

*For more about Sarah, see
www.healthtalkonline.org/medical_research/clinical_trials/Topic/3638/Interview/2017/Clip/14719*

*For interviews with other people who considered withdrawing from a clinical study, see
www.healthtalkonline.org/medical_research/clinical_trials/Topic/3638*



RESEARCH STORIES: Bias in Last Observation Carried Forward Method

For some conditions, such as dementia, patients typically worsen in their cognitive functioning over time. That means that a patient assessment collected midway through a trial will overestimate cognitive functioning at the end of the trial. If we want to understand a patient's cognitive functioning at the end of a trial, 10 months after starting a therapy, we cannot assume that earlier assessments (e.g., at 6 months) of patients who dropped out of a trial can be "carried forward" to the end of the trial as a substitute for the final planned assessment.

The figure below illustrates the bias that results from an imputation method called the Last Observation Carried Forward (LOCF) method, which has been a common solution to the problem of patients dropping out of trials before their final planned visit. Consider a patient randomized to the control treatment (line b) who drops out of the trial soon after his 6-month assessment. If the trial investigators simply substitute this assessment for the planned final assessment, they will overestimate his level of cognitive functioning at the end of the trial. The difference between the assessed value at 6 months and the true value at 10 months is shown in the figure as the LOCF bias (Molnar et al. 2009).

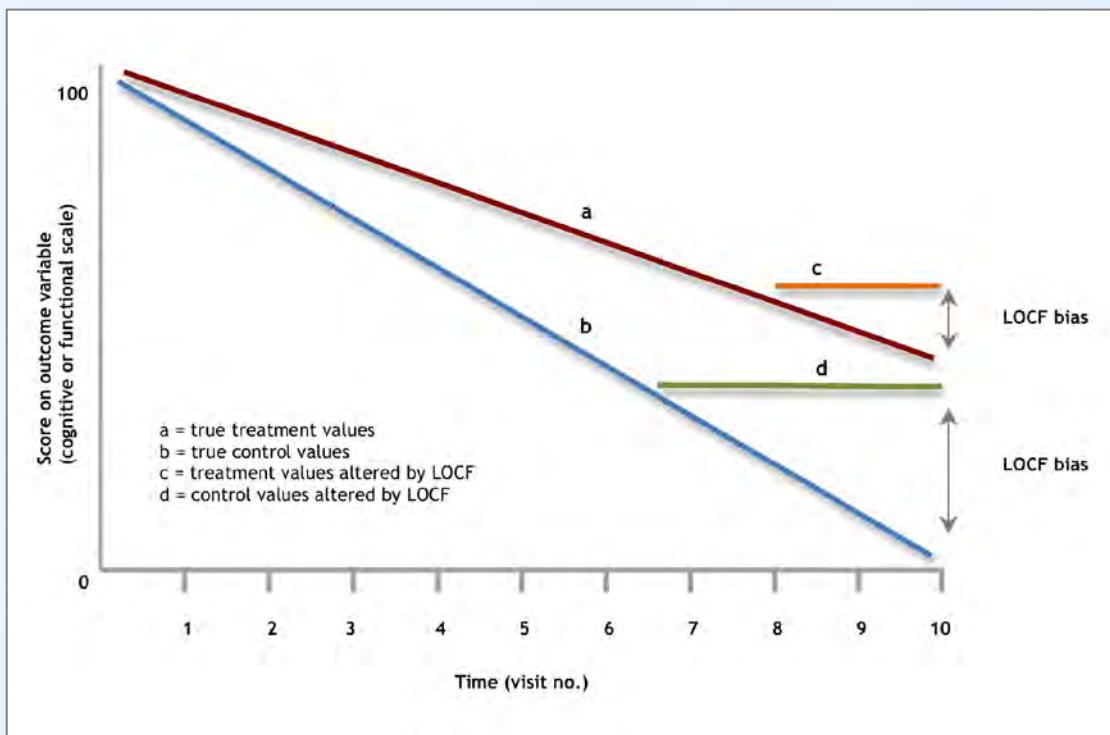


Figure from Molnar (2009) reprinted under the Creative Commons Attribution Share Alike License. Any derivative use of this work must be distributed only under a license identical to this one and must be attributed to the authors. The authors retain copyright of their work.

Cross-Cutting Standards for PCOR

4: Standards for Preventing and Handling Missing Data

MD-1 Describe methods to prevent and monitor missing data

Investigators should explicitly anticipate potential problems of missing data. The study protocol should contain a section that addresses missing data issues and steps taken in study design and conduct to monitor and limit the impact of missing data. Missingness can occur from patient dropout, failure to provide data, and/or administrative or data management issues. As relevant, the protocol should include the anticipated amount of and reasons for missing data, as well as plans to follow up with participants. This standard applies to all study designs for any type of research question.

MD-2 Describe statistical methods to handle missing data

Statistical methods for handling missing data should be pre-specified in study protocols. The reasons for missing data should be considered in the analysis. The plausibility of the assumptions associated with the approach should be assessed. A discussion of the potential ramifications of the approach to missing data on the results should be provided. This standard applies to all study designs for any type of research question.

MD-3 Use validated methods to deal with missing data that properly account for statistical uncertainty due to missingness

Statistical inference of intervention effects or measures of association should account for statistical uncertainty attributable to missing data. This means that methods used for imputing missing data should have valid Type I error rates and that confidence intervals should have the nominal coverage properties. This standard applies to all study designs for any type of research question. Bayesian methods and methods such as multiple imputation satisfy this condition, along with various likelihood-based and other validated methods. Single imputation methods like last observation carried forward and baseline observation carried forward are discouraged as the primary approach for handling missing data in the analysis. If investigators do use single-based imputation methods, they must provide a compelling scientific rationale as to why the method is appropriate.

MD-4 Record and report all reasons for dropout and missing data, and account for all patients in reports

Whenever a participant drops out of a research study, the investigator should document the following: 1) the specific reason for dropout, in as much detail as possible; 2) who decided that the participant would drop out; and 3) whether the dropout involves some or all types of participation. Investigators should attempt to continue to collect information on key outcomes for participants unless consent is withdrawn. This standard applies to all prospective study designs that aim to assess intervention effectiveness. All participants included in the study should be accounted for in the report, whether or not they are included in the analysis. Describe and justify any planned reasons for excluding participants from analysis.

4: Standards for Preventing and Handling Missing Data (Continued)

MD-5 Examine sensitivity of inferences to missing data methods and assumptions, and incorporate into interpretation

Examining sensitivity to the assumptions about the missing data mechanism (i.e., sensitivity analysis) should be a mandatory component of the study protocol, analysis, and reporting. This standard applies to all study designs for any type of research question. Statistical summaries should be used to describe missing data in studies, including a comparison of baseline characteristics of units (e.g., patients, questions, or clinics) with and without missing data. These quantitative results should be incorporated into the interpretation of the study and reflected in the discussion section and possibly the abstract.

5. Standards for Heterogeneity of Treatment Effect (HTE)

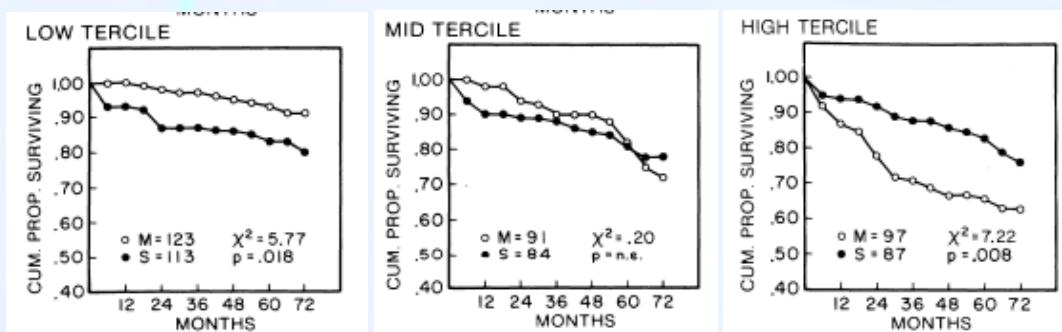
Heterogeneity of treatment effect (HTE) is a technical term for the fact that different people do not always respond the same way to the same treatment. In some, the treatment will produce the intended benefit; in others, the benefit may be less than what was intended. And yet in others, the treatment may have no effect at all or may even cause harm. In clinical research, this variability can be masked by the study design and analysis, by not measuring the variables that predict different responses, or by not analyzing these variables. In both clinical trials and observational studies, results are often averaged across all the patients in a study, obscuring how responses to a treatment might vary across individuals within the study population. As a result, it can be hard to determine from research results what the effect of a treatment will be for a specific type of patient.

Explicitly addressing HTE in research helps answer the question, “What is likely to happen to patients like me?” This makes research results more useful for patients and clinicians who need to decide the best course of treatment (see **Research Stories: Heterogeneity of Treatment Effects**).



RESEARCH STORIES: Heterogeneity of Treatment Effects

The figures below show six-year survival rates during the 1970s for patients with chest pain (angina) at high risk for mortality from heart disease. Patients were randomly assigned to heart bypass surgery (black dots) or a non-surgical treatment (white dots). The three panels depict patients at high, medium, and low risk for mortality. The risk categories were determined by four noninvasive factors: electrocardiogram (ECG or EKG) results, presence of hypertension, a previous instance of heart attack, and a marked limitation in the patient's ability to perform everyday activities without difficulty (e.g. pain, shortness of breath, dizziness). The figure shows that the best treatment differed for patients depending on their risk of mortality before starting treatment (Detre et al. 1981). A low-risk patient (with a normal EKG? and no history of heart attack or high blood pressure, who is able to perform everyday activities without strain) would live longer without an invasive bypass surgical procedure, while those patients at high risk (with an abnormal EKG and/or history of high blood pressure or previous heart attack, who cannot function normally in everyday activities) would live longer if treated with bypass surgery. Consequently, the most appropriate treatment for chest pain is heterogeneous (varies) across patients. Treatments for patients with angina have improved since the early 1970s, but the statistical approach to evaluating treatment effects and how they depend on patient characteristics remains useful today (Sox and Goodman 2012).



Figures from Detre (1981) reprinted by permission of Wolters Kluwer Health provided by Copyright Clearance Center.

Analysis that focuses on HTE could include either 1) an estimation of separate treatment effects for subgroups of patients or 2) predictions of whether a specific person will benefit from treatment. The most common approach is to use subgroup analyses to estimate the effects of treatments in a specified subset of the study patients. Currently, predicting individual effects occurs less often, though it is of increasing interest as the field of personalized medicine grows and such tools as decision

analysis and microsimulation models are developed to predict how patients will benefit. However, in this initial group of standards, those for HTE research apply only to estimating subgroup treatment effects.

Researchers often estimate the effect of treatment separately for patient groups by stratifying by subgroup (e.g., men versus women). This approach is susceptible to the well-known problem of multiple post hoc analyses that can yield an increased likelihood of falsely detecting HTE (referred to in statistics as Type I error) or failing to detect true HTE (Type II error). Although estimating stratified treatment effects may be valid for testing a limited number of subgroups (when sample sizes are large enough), this approach is incorrect for inferring HTE when multiple subgroup comparisons are required.

Rationale for the Standards

The first step in assuring high-quality HTE analyses is understanding the purpose of the research. Therefore, the standards require that researchers state their goal (**HT-1**). This statement will direct the appropriate design and analysis plan for the study and also allow stakeholders to interpret results correctly. Next, specifying subgroups and reporting the number of subgroups tested ensures that methods are transparent and that errors from multiple statistical comparisons (e.g., Type I or II errors) are detected or avoided (Goldfine et al. 2011; Lagakos 2006; Brookes et al. 2001) (**HT-2**).

Once a study is designed and conducted, no matter the purpose, testing for HTE involves determining via an interaction test whether the difference between the treatment effects for the subgroups is zero. This requirement applies in both randomized trials and observational studies. In the former, patients are randomized to the intervention; subgroups are not randomized. Thus, the subgroups may have different baseline characteristics, which confound the interpretation of results. Regardless of the type of study, interaction tests require fewer comparisons and have more statistical power than direct subgroup analyses. In some cases, the use of multiple analytic methods to look for consistent effects, while accounting for the different limitations of all the methods, may be the most useful strategy for drawing valid conclusions (**HT-3** and **HT-4**).

Cross-Cutting Standards for PCOR

5: Standards for Heterogeneity of Treatment Effects

HT-1 State the goals of HTE analyses

State the inferential goal of each HTE analysis, specifying how it is related to the topic of the research, translate this into an analytic approach, and highlight the linkages between the two. Identify analyses as hypothesis driven (sometimes denoted confirmatory), or hypothesis generating (sometime denoted exploratory).

HT-2 For all HTE analyses, pre-specify the analysis plan; for hypothesis-driven HTE analyses, pre-specify hypotheses and supporting evidence base

The study protocol should unambiguously pre-specify planned HTE analyses. Pre-specification of hypothesis-driven HTE analyses should include a clear statement of the hypotheses the study will evaluate, including how groups will be defined (e.g., by multivariate score or stratification) and outcome measures, and the direction of the expected treatment effects. The pre-specified hypotheses should be based on prior evidence, which should be described clearly in the study protocol and published paper.

HT-3 All HTE claims must be based on appropriate statistical contrasts among groups being compared, such as interaction tests or estimates of differences in treatment effect

A common error in HTE analyses is to claim differences in treatment effect when one group shows a statistically significant treatment effect and another does not. To claim differences in treatment effect among subgroups, appropriate statistical methods must be used to directly contrast them. Such contrasts include, but are not limited to, interaction tests, differences in treatment effect estimates with standard errors, or a variety of approaches to adjusting the estimated subgroup effect, such as Bayesian shrinkage estimates. Within each subgroup level, studies should present the treatment effect estimates and measures of variability.

HT-4 For any HTE analysis, report all pre-specified analyses and, at minimum, the number of post hoc analyses, including all subgroups and outcomes analyzed

Protocols and study reports must report the exact procedures used to explore HTE, including data mining or any automatic regression approaches. HTE analyses should clearly report the procedures by which subgroups were defined, (e.g., by categorical predictors or continuous risk scores), and the effective number of subgroups and outcomes examined. If a non-prespecified stratum or subgroup is claimed to show a treatment effect that is different from others, methods should be used that account for the number of contrasts examined. These methods include, but are not limited to, p-value adjustment, false discovery rates, Bayesian shrinkage estimates, adjusted confidence intervals, or validation methods (internal or external).

Standards for Specific Study Designs and Methods

6. Standards for Data Registries

A registry is an organized system that collects data for scientific, clinical, or policy purposes and can provide data for observational studies. Clinical registries are structured systems for collecting and organizing uniform data about the progress and outcomes associated either with the course of a disease or with the defining characteristic of the patients (e.g., familial cancer risk or device implantation).

Registries may compile data from different sources, such as medical records and lab reports, or across multiple healthcare settings, such as all hospitals in a state or all hospitals and physicians' offices in a region. They also can be a way to prompt or require the collection of additional data about a group of patients with a specific condition (e.g., diabetes or cancer) who undergo a diagnostic test (e.g., a PET scan) or have a particular treatment (e.g., hip replacement). For example, a cancer registry could include information from medical charts, surgery reports, and tumor pathology studies and then prompt clinicians to collect information on patients' symptoms using a standardized questionnaire.

Registries have led to significant discoveries about the comparative effectiveness of treatments. Collecting post-operative data about a group of patients who had hip replacement, for example, allowed researchers to uncover a significant problem with one type of artificial hip (see **Research Stories: National Joint Registry of England and Wales**).



RESEARCH STORIES: National Joint Registry of England and Wales

The National Joint Registry of England and Wales, the world's largest registry of hip replacements, contains records of more than 400,000 first, or "primary," hip replacements. It tracks hip replacements performed since 2003 and documents when the joints fail, requiring patients to undergo a second surgery. The size of the registry allowed orthopedic surgeons and other investigators to compare the effectiveness of different materials used in the replacements, and thereby discover a fault much more quickly than if they had relied on patient reports in regular practice. The registry data showed that metal-on-metal hip replacements are more likely to fail in the five years after hip surgery than metal-on-ceramic or ceramic-on-ceramic products.

A 60-year-old man undergoing a primary hip replacement with a relatively small (28-millimeter-diameter) ceramic-on-ceramic product can expect a 2.0% risk of product failure during the first five years, while the same man with a similar metal-on-metal product can expect a 3.2% risk of product failure.

The registry's 31,171 records of patients with metal-on-metal implants enabled the investigators to determine that the failure rate increased with the diameter of the implants—especially in younger women. The registry was also large enough to demonstrate that the higher failure rate could not be explained by a single manufacturer's product, so appears to be a problem for all metal-on-metal implants. The orthopedic surgeons analyzing the registry data recommended against future hip replacements with metal-on-metal devices and suggested annual review of patients who already had these implants (Smith et al. 2012).

Patients are usually included in registries in anticipation of future research related to the focus of the registry. When questions arise that can be answered with such data, answers can often be obtained quickly because of the comprehensiveness of the risk and outcome data already in the registry. Registries are particularly important for PCOR. When properly designed, they can provide data on groups of patients not always included in clinical trials, and they can be very responsive to rapid changes in medical practice. Registries can also be used to study factors that are difficult or impossible to randomize, such as clinician or patient behaviors, and factors that predict who is more likely to experience the benefits or harms of different treatments (see **Research in Practice: Data Registries**). The fact that registries are based on medical care as it is actually delivered in real-world situations increases the likelihood that the findings will be broadly applicable to many people and situations.



RESEARCH IN PRACTICE: Data Registries

Jacqueline Fridge, MD, is a pediatric gastroenterologist in Portland, Oregon. Two years ago she led her practice, Northwest Pediatric Gastroenterology LLC, to join the ImproveCareNow collaborative, a national health network that uses collaboration and data to drive improvements in the care and health of children with Crohn's disease and ulcerative colitis (Crandall 2009).

How has the use of a registry affected your practice?

Jacqueline Fridge: To a certain degree it's standardizing care between physicians. We have not yet done a lot of physician-to-physician comparison, but that is the next step, especially when you are looking at remission rate, we're going to want to see if there is an outlier. And then drill down to see if there are differences. What practices does that physician have? Do they have a genuinely more challenging group of patients for some reason or is their practice different than ours?

For example, are their procedures not being performed correctly or are they being performed in a different way?

Fridge: Right, or are they not getting the labs as often as ours? Who knows, maybe I'm the outlier. So, I think that's kind of the way registries are impacting our care.

Have you used registries to answer patient questions?

Fridge: One of the things ImproveCareNow is doing, because they have such a huge number of patients, is looking at some of the trials that were previously done. They can look through their research data and see if, in real life, the outcomes replicate the study. They replicated REACH, which is one of the original Infliximab (Remicade®) studies [this drug treats rheumatoid arthritis, psoriatic arthritis, ankylosing spondylitis, Crohn's disease, plaque psoriasis, and ulcerative colitis], and by pulling the data out of the ImproveCareNow database, they showed that the results almost exactly matched REACH. So I think more of that type of data reinforcement is going to be coming down the road and I think it is going to be able to help answer questions.

Have registries provided any particular education or expertise about the course of inflammatory bowel disease that might not have come to light otherwise?

Fridge: I think what ImproveCareNow is giving us is a volume of data that we've never had before. The registry is much more proactive, it's not just this data-collecting machine. Each month they say, "What are you testing this month, what quality improvement are you working on currently?"

I think what the registry is going to do is formalize a lot of anecdotal thinking. An example is the Cystic Fibrosis Foundation and cystic fibrosis registries. They started off with a registry, then they had the Improve Cystic Fibrosis centers, each one funneling data and information into the registry, and then they took some of those centers and made them the test centers for their drug trials. So I think there's very much a hope and expectation that we'll actually start to get pediatric data.

But the same characteristics of registries that make them reflective of real-world practice sometimes limit their usefulness in informing healthcare decisions. Data derived from clinical sources may not be as tightly controlled as data collected in a clinical trial or even some prospective cohort studies, and definitions may differ across data sources and change over time. This is where methodological standards are useful. If the potential of registries is to be realized, careful planning is needed prior to establishing a registry, and researchers designing studies based on registries need to understand the data and be sure of its quality and relevance for their study. Furthermore, registry data analysis needs to formally consider other factors (referred to as confounding factors) that might influence the results. Well-constructed, well-implemented registry studies can promote patient-centeredness by providing timely data pertinent to clinician and patient decision making, but to do so registries need to contain relevant, high-quality data that is used appropriately.

Rationale for the Standards

The quality of data derived from registries depends on a wide array of factors, including design, data elements, data sources, governance, and maintenance. Without careful planning and oversight, there can be problems with the use and confidentiality of registry data. Also, tracking and matching patients across data sources and over time is resource-intensive and must be planned carefully to avoid confidentiality breaches. Because registries typically follow the natural history of patients, they require multiple points of follow-up. Registries are often most useful when they are maintained over periods that are long enough to provide important long-term outcomes that are important to patients (see **Patient Voices: Suzanne**). However, the problem of missing data may be significant in studies based on registries requiring long-term data collection that includes multiple patient contacts.

Standard **DR-1** specifically addresses the construction and maintenance of the registry. Registries are most likely to generate usable inferences if their construction is based on a protocol related to at least one clinical question and includes plans for enrollment, patient follow-up, and data linkage. Such protocols must also include details of consent procedures and confidentiality protections that take into account the possibility of re-identification. Planning how best to collect and aggregate the data, protect patient privacy, document changes, and ensure data quality increases the likelihood that the registry can answer essential PCOR questions.



PATIENT VOICES: Suzanne

Suzanne has had juvenile-onset rheumatoid arthritis for 22 years.

I've had both knees replaced, and the surgery and the rehabilitation occurred just as I expected and just as I'd been told. There were no surprises because of the large body of evidence (e.g., research, knowledge of the rheumatology provider community) about the results of knee surgery. Eight years after my knee replacements, it came time to tackle my wrists. Several of the small bones in my right wrist had grown together, preventing any significant movement. In other places in my right wrist, the bone had eroded. The bones in my right wrist were so badly damaged that the surgeon could flake pieces off of bone with his thumb.

Wrist replacement was now not an option, and a total fusion of the joint—removing all of the soft tissue and inserting some hardware to compel the bones to finish growing together—was the best way to alleviate pain and restore function. With this option, though, the hand would forever extend in a straight line from the forearm; no bending, no twisting, and no turning. None of the arthritis patients I know had gone through a wrist fusion or a wrist replacement—at least not within the past 10 years. While the surgery team was excellent and provided ample information on the procedure itself, I was not aware of any registries or much research about patients' views on the outcomes of this surgery.

I opted to move forward with the surgery, fingers crossed. If the only goal was to alleviate pain in the right wrist, the surgery was a complete success. Four years after the surgery, my right wrist was one of my best joints—strong, sturdy, and pain-free. What I did not expect was the effect of the surgery on my right hand and fingers. Now that the wrist isn't mobile, the fourth and fifth fingers and the fourth and fifth metacarpal phalangeal joints on that hand have picked up much of the slack. The added stress to these areas has led to new joint deformities and challenges. Was it worth it? It is hard to say. The wrist pain and instability were significant functional issues, but I wonder if there were other options that could have fixed the wrist and not exacerbated the arthritis in the hand and fingers.

Now, I need to focus on whether I should have wrist replacement surgery or have a wrist fusion on the left wrist. Will a wrist replacement work for me? What will be the effect of wrist replacement on the fingers and hands? If I opt for a fusion instead, is there a way to preserve the fingers and hand or should I expect the same functional impact as with the right wrist? Are there other surgical options beyond these two?

Before I launch into another surgery with unintended consequences, I would really like to see information about how other people with my condition have responded to wrist surgery and what my best options are, but as of now, I am not aware of any available information.

The other standards on data registries apply to researchers who conduct studies using data derived from registries. Researchers need to consider the same elements of the registry that were considered when it was designed—as well as the advantages and limitations of the registry’s data for their particular research question (**DR-2**). Researchers also have to pay attention to issues of data quality and potential biases in studies that utilize registry data because registries may not gather all the information needed for certain questions that arise after the registry is established, can be affected by a variety of time trends, and do not always include control populations (i.e., patients who do not receive treatment).

Perhaps the chief consideration is that risk factors for the outcomes may not be equally distributed among the groups of patients being compared. This problem, known as confounding, occurs because patients are not randomized and may be included in the registry for a variety of reasons. Research based on registries must contain data elements that will allow for statistical controls for confounding, and researchers must develop complementary approaches for their data analyses (**DR-3**). Wrong conclusions can be drawn if the data collected are not standardized (e.g., in definitions or follow-up); large amounts of data are missing; or confounding is not controlled.

Standards for Specific Study Designs and Methods

6: Standards for Data Registries

DR-1 Requirements for the design and features of registries

Registries established for conducting PCOR must have the following characteristics to facilitate the collection and aggregation of usable data, to ensure appropriate privacy and confidentiality, to document changes to the registry protocol, and to guide robust analyses that include important confounders.

A. Patient Follow-up

The objective(s) of the registry should determine the type, extent, and length of patient follow-up. Describe what triggers the follow-up, the follow-up measures, and the last contact with the patient. Ensure that the planned follow-up time is adequate to address the main objective and that planned patient-retention efforts are suitable to the target population and anticipated challenges. Describe expected loss to follow-up and potential effect on the results, including possible biases resulting from differential loss.

B. Data Safety and Security

Registry custodians should provide transparency for institutional review boards by describing data use agreements, informed consent, data security, and approaches to protecting security including risk of re-identification of patients. If using previously collected data, describe how these address the risk of re-identification of patients and the actual use of data compared with the originally designed and consented use of the data.

C. Data Quality Assurance

A quality assurance plan for registries should address: 1) structured training tools for data abstractors; 2) use of data quality checks for ranges and logical consistency for key exposure and outcome variables and covariates; and 3) data review and verification procedures, including source data verification plans and validation statistics focused on the key exposure and outcome variables and covariates for which sites may be especially challenged. A risk-based approach to quality assurance is advisable, focused on variables of greatest importance.

D. Document and Explain Any Modifications to the Protocol

Modifications to a registry protocol may be necessary for a variety of reasons. When modifications are necessary, they should be explained, documented, and made available to anyone planning to use the registry data.

E. Consistent Data Collection

Clear, operational definitions of data elements should be provided. Create and distribute standard instructions to data collectors. Use standardized data element definitions and/or data dictionaries whenever possible. When creating a new registry, published literature should be reviewed to identify existing, widely used definitions before drafting new definitions.

F. Systematic Patient Enrollment and Follow-up

Enroll patients systematically and follow them in as unbiased a manner as possible, using similar procedures at all participating sites. Describe how patients and providers were recruited into the study to allow the impact of selection bias to be clearly understood; for example, by explaining whether the sampling was population-based or otherwise and any efforts employed to confirm the quality of adherence to agreed-on enrollment practices.

6: Standards for Data Registries (Continued)

DR-1 Requirements for the design and features of registries (Continued)

G. Monitor and Minimize Loss to Follow-up

Monitor loss to follow-up to ensure that follow-up is reasonably complete for the main objective. Minimizing loss to follow-up requires having a target and advance planning for what actions will be employed in the event that this target is in jeopardy. At the outset of the registry, develop a patient retention plan that documents when a patient will be considered lost to follow-up and what actions will be taken to minimize such loss. At the enrollment visit, consider collecting multiple types of contact information (e.g., telephone, mailing address, and email address) for the patient, as well as collecting contact information for an alternate contact if the patient cannot be reached directly. Verify contact information at each subsequent visit and update as needed. When a patient misses a visit, contact the patient following a standard protocol (e.g., phone call one day after missed visit, email one week after missed visit). If the patient withdraws from the registry, attempt to document the reason for withdrawal so that issues can be identified and addressed (e.g., overly burdensome patient-reported outcome measures). Efforts at minimizing loss to follow-up should be tempered by considerations and sensitivity to repeated intrusions on patients and to the health conditions and interventions under study. Consider collecting enough information to permit accurate linkage with other data sources, such as the National Death Index, for long-term follow-up.

H. Collect Data to Address Confounding

Registries should identify important potential confounders during the planning phase and collect reasonably sufficient data on these potential confounders to facilitate the use of appropriate statistical techniques during the analysis phase.

DR-2 Selection and use of registries

Researchers planning PCOR studies relying on registries must ensure that these meet the requirements contained in Standard DR-1 and must document each required feature of the registry(s) to be used (e.g., in an appendix to the funding application or study protocol). Deviations from the requirements should be justified by explaining why a required feature is not feasible or not necessary to achieve the overall goals of Standard DR-1.

DR-3 Robust analysis of confounding factors

In studies that use registries to evaluate the comparative effectiveness or safety of interventions, investigators should select an approach for adjusting for known and measured confounders, such as multivariable regression analysis or propensity scores to create matched comparison groups, or an instrumental variable analysis if a valid instrument is available. It is also desirable to examine the robustness of the results through sensitivity analyses focused on testing key assumptions and evaluating the likely impact of unmeasured confounders. The rationale for using selected techniques, any assumptions made, and the strengths and limitations of the techniques should be described in reports of the study findings to allow for informed interpretation of the results.

7. Standards for Data Networks as Research-Facilitating Infrastructures

Collaborative data networks are agreements that coordinate data use across healthcare organizations. Data networks can improve delivery of clinical services and augment healthcare research. Data networks can aggregate information from a range of data sources (e.g., claims, medical records, lab/pathology reports) or from various locations (e.g., health plans, hospitals, clinics, care facilities). The infrastructure created by a network may then be used to establish disease-specific registries, maintain broad-ranging surveillance systems, or facilitate the conduct of randomized trials. Data networks may cover a wide range of research topics, including but not limited to studying the effectiveness of diagnostic tests, monitoring adverse effects of new drugs or devices, and testing new cancer treatments.

Data networks that facilitate research include such key components as a data architecture (structure), privacy policies that protect patient information, governance guidelines that specify roles and responsibilities, and rules for how data elements are defined, described, and organized. But these components do not determine research questions for clinical studies or research design.

Data networks have many characteristics that make them important for the development and advancement of PCOR. Analyzing data already collected across organizations or locations is more efficient than replicating studies in multiple locations or populations. Studies based on networked data are also likely to include more types of patients and variations in treatment patterns than would be available in any one site. This variety means the results are more likely to be generalizable—useful to more patients and clinicians when they have to make decisions. Almost by definition, data networks include larger numbers of patients than can be enrolled in most trials and cohort studies. While a larger number of patients alone does not necessarily improve a study, it can increase the precision in effect estimates and make it possible to detect smaller differences in outcomes or recognize differences in a shorter period. With large numbers of records, it is easier to determine whether the comparative effectiveness varies across meaningful subgroups (e.g., between men and women or among people with different comorbidities).

Despite all these advantages, a data network is only as good as the quality of its data. The challenges in establishing and maintaining data networks include harmonizing both the technical aspects and the expectations and responsibilities of the participating organizations. Definitions and other

characteristics of data elements need to be clear, agreed upon, and verified. Creating and maintaining standardized terminology and data descriptions require planning and resources. Also needed are agreement and clarity about how patient privacy will be protected, who has access to the data, and who owns both the data and the research results. Setting standards for data networks ensures that key components are included when networks are designed and that these components are also considered when data from these networks are used in research studies.

Rationale for the Standards

Several organizations in the United States, Canada, and Europe have developed guidelines, best practices, and initiatives for defining crucial characteristics of data networks. These range from specific projects to standardize terminology, to recommended models for network structures, to laws or policies that are specific to health care—like the Health Insurance Portability and Accountability Act (HIPAA)—or general policies with applications in health care, such as the Organization for Economic Co-operation and Development personal privacy guidelines (OECD 2013). Compiling all of these would result in a lengthy technical manual that could be prescriptive and thereby hamper innovation in what is a rapidly evolving and growing field. The standards for data networks will help to ensure data quality, privacy, and collaboration. These standards are intended to apply to networks that supply data for comparative effectiveness research (CER) or PCOR. These standards are not proposed for network initiatives that have other purposes.

For a data network to function and provide useful data, processes have to be created and documented that transform data elements so they are equivalent even when they come from different sources. Data networks link and share information about individuals in ways that could compromise patient privacy. Generally, study proposals and protocols should describe data use agreements, informed consent, data security, and approaches to protecting security. Proposals should also describe how these address the risk of re-identifying patients and the actual use of data compared with the originally designed and consented use. For patients and clinicians to realize the benefits of research via data networks without jeopardizing privacy, standards are required to limit and control who has access to the data. Additionally, data networks need to evaluate proactively whether any use or structural characteristic of the network is likely to compromise confidentiality.

The usefulness of a data network often increases with the longevity of the network. Longevity requires that the participating organizations maintain relationships and continue to collaborate. These relationships can be complex, and the agreements are often detailed and cover a range of roles and responsibilities. At a minimum, agreement needs to exist about ownership of both the data and the products resulting from the network (intellectual property policies). Another important aspect is the need for standardized terminology, and information (known as metadata) about the data elements must be provided. Then, data elements also should assembled into a model that shows the relationships among the data elements and helps all users to interpret the data correctly.

The PCORI standards for data networks recognize that the construction and management of the network is separate from the use of data for CER. The first standard addresses development and maintenance of a network's policies and procedures (**DN-1**). The second standard addresses the activities of researchers who seek to access and use data from an existing network (**DN-2**).

Standards for Specific Study Designs and Methods

7. Standards for Data Networks as Research-Facilitating Structures

DN-1 Requirements for the design and features of data networks

Data networks established for conducting PCOR must have the following characteristics to facilitate valid, useable data and to ensure appropriate privacy, confidentiality, and intellectual property protections:

- A. Data Integration Strategy**—In order for equivalent data elements from different sources to be harmonized (treated as equivalent), processes should be created and documented that either 1) transform and standardize data elements prior to analysis or 2) make transformation logic available that can be executed when data are extracted. The selected approach should be based on an understanding of the research domain of interest.
- B. Risk Assessment Strategy**—Data custodians should measure the risk of re-identification of data and apply algorithms to ensure that the desired level of confidentiality is achieved to meet the need of the particular PCOR application.
- C. Identity Management and Authentication of Individual Researchers**—Develop reliable processes for verifying credentials of researchers who are granted access to a distributed research network and for authenticating them.
- D. Intellectual Property Policies**—A research network should develop policies for the handling and dissemination of intellectual property (IP); networks should also have an ongoing process for reviewing and refreshing those policies. IP can include data, research databases, papers, reports, patents, and/or products resulting from research using the network. Guidelines should balance 1) minimizing impediments to innovation in research processes and 2) making the results of research widely accessible, particularly to the people who need them the most.
- E. Standardized Terminology Encoding of Data Content**—The data contents should be represented with standardized terminology systems to ensure that their meaning is unambiguously and consistently understood by parties using the data.
- F. Metadata Annotation of Data Content**—Semantic and administrative aspects of data contents should be annotated with a set of metadata items. Metadata annotation helps to correctly identify the intended meaning of a data element and facilitates an automated compatibility check among data elements.
- G. Common Data Model**—Individual data items should be assembled into a contextual environment that shows close or distant association among data. A common data model (CDM) specifies necessary data items that need to be collected and shared across participating institutes, clearly represents these associations and relationships among data elements, and promotes correct interpretation of the data content.

DN-2 Selection and use of data networks

Researchers planning PCOR studies relying on data networks must ensure that these networks meet the requirements contained in Standard DN-1, and they must document each required feature of the data network(s) to be used (e.g., in an appendix to the funding application or study protocol). Deviations from the requirements should be justified by explaining why a required feature is not feasible or not necessary to achieve the overall goals of standard DN-1.

8. Standards for Causal Inference Methods

One of health research's key objectives is to determine the causes of a health outcome. This is the information that patients, families, and clinicians most frequently want—will the treatment they choose cause improvement in the outcomes they care about?

The challenge is that when the “cause” is a medical intervention or treatment, it can be difficult to separate the effects of the treatment from other factors that might vary between patients who had the treatment and those who did not. Randomized controlled trials (RCTs) are a methodological answer to this problem. Because they randomly assign participants to a treatment, the distribution of risk factors for the health outcome—known as potential “confounders” of the causal relationship—is likely to be similar across the groups under review. Consequently, on average, across all the different possible assignments of patients, the estimate of how much the intervention affects the outcome would be correct, even if individual participants differ in ways besides the treatment they receive.

However, randomization cannot solve all of the challenges in handling confounders and identifying whether the treatment is the cause of the outcome. For some settings and questions, a randomized trial is impossible, undesirable, unethical, and/or would require too many resources. In these circumstances, researchers use observational methods—study designs in which the interventions are decided not by random assignment but as part of the normal process of clinical care. The challenge is that the complexity and variability of patients and their circumstances, as well as the care they receive, often make it difficult to conclude whether a specific treatment is responsible for the observed clinical outcomes.

Several analytic tools are available that aim to control mathematically the effects of confounding variables and thereby produce a valid estimate of a treatment's effect even in complex situations. These methods are varied. Some are well established; others are promising but still developing. While these tools are both powerful and useful, they have drawbacks—most notably, the majority of methods can control only for the effect of confounders that were actually identified. A subtler issue is that the methods were not founded on a clearly articulated definition of “cause,” so that they cannot distinguish between something that truly causes an outcome and something that merely happens at the same time.

The analytic tools designed to address confounding and approach an understanding of causality are called “causal inference methods.” The methods include various forms of population restriction and regression methods. Some, such as instrumental variable methods, were adapted from other fields, in this case, economics, where they have used them for decades. Others, such as propensity scores, have seen wide use in biomedicine only over the past decade, even though they were developed earlier.

The various methods also address the issue of confounding differently. Like standard regression methods, propensity scores cannot solve the problem of unmeasured confounding factors, but they can adjust for multiple confounders and variables that serve as proxies for unmeasured confounders (Rosenbaum and Rubin 1984). Instrumental variable methods, on the other hand, purport to get around the unmeasured confounder problem by identifying and exploiting naturally occurring distributions of treatment choices in the healthcare system that resemble randomization but often rely on assumptions that are untestable using the data available. Thus, these assumptions require extraordinarily close scrutiny.

None of these methods solve the problems of causal inference posed by observational studies, but they can produce more accurate estimates of effect and uncertainty. Researchers using causal inference methods try to set up the analysis of observational data to be as much like a randomized trial as possible and to decrease the chance of reaching spurious, incorrect conclusions (see **Research Stories: Human Immunodeficiency Virus**).

Although the literature about causal inference and the development of related analytic methods are in their early stages, they are being used with increasing frequency in PCOR. One reason is that electronic healthcare databases make it possible to conduct observational studies in large populations in standard clinical settings. As few of those studies are randomized, observational methods are being increasingly relied upon to extract estimates that can support causal interpretations from data that may not have been produced or gathered with such an intended use.



RESEARCH STORIES: Human Immunodeficiency Virus

What is the optimal time for patients with HIV infection to start combined antiretroviral therapy? Strategic Timing of Anti-Retroviral Therapy (START), a randomized controlled trial that addresses that question, will be completed in 2016. In the interim, investigators from the HIV-CAUSAL Collaboration conducted a comparative effectiveness cohort study in 20,971 patients. The team used advance statistical methods—called dynamic marginal structural models—that improved its capacity to emulate randomized controlled trials by correcting for changes in treatment and health status over time. Conventional statistical methods may generate biased findings when physicians change treatment in response to changes in patient health, so marginal structural models mark a major advance for studies in which patients are not assigned randomly to different treatment strategies.

Using routine healthcare data from the Veterans Health Administration and HIV clinics in Europe, the investigators considered the question of whether to start combined antiretroviral therapy earlier (before the laboratory measure of immune function drops below a relatively high threshold) or later (after the measure drops below an intermediate or lower threshold). The marginal structural model revealed that starting treatment earlier is more effective at reducing the rate of mortality and AIDS-defining illness (the diseases associated with AIDS). Patients who delayed starting this therapy until the low laboratory threshold suffered a 38 percent increase in the rate of mortality and AIDS-defining illness (Hernán and McKaig 2011).

Rationale for the Standards

In all studies, researchers need to pay substantial attention to possible sources of bias and address biases in order for the results to produce valid conclusions about whether a treatment causes an outcome. Exactly how bias is avoided or addressed varies according to the study design and the research question. One approach used in observational studies is to choose patients based on a group of variables, referred to as a covariate history, that is known at the start point and not based on later changes (**CI-1**). Measuring and adjusting for pretreatment variables is common in observational studies and is an acceptable approach for mimicking randomization at baseline. However, if these variables are measured again (or if adjustments are made based on those variables) between baseline and follow-up, then researchers may introduce bias if these variables are affected by the study treatment. Such bias may make it harder to ascertain whether the treatment is causing the result.

Regardless of the type of study design, an obvious starting point is to specify who is included in the analysis, why they are included, whether any variables measured after baseline may introduce bias, and how the different groups compare on key characteristics (**CI-2**). To increase the accuracy of results, researchers may include only selected patients in some analyses. For example, patients might be separated by age or the severity of their illness. In some cases, statistical methods, such as propensity scores, can combine several characteristics into one variable that is used to control for baseline differences.

In studies where the treatment is not controlled by the researchers, it is also important that timing of the outcome measurement relative to the treatment or exposure is clearly defined (**CI-3**). When this timing is known and logical, it can help verify the potential causal relationship. Similarly, variables considered confounders should be measured before the treatment. If these variables change over time, this change needs to be addressed in the study design or analysis (**CI-4**).

Creating standards specific to all current causal inference methods was not feasible at this time. In this initial group, standards are included for propensity scores (**CI-5**) and instrumental variables (**CI-6**), as these are relatively well-developed methods that are increasingly used in PCOR. When sophisticated analytical approaches are used, transparency is particularly important. These standards specify that additional efforts are required to document the assumptions underlying the analyses and how these assumptions were tested.

Standards for Specific Study Designs and Methods

8. Standards for Causal Inference Methods

- CI-1 Define analysis population using covariate histories**
- Decisions about whether patients are included in an analysis should be based on information available at each patient's time of study entry in prospective studies or on information from a defined time period prior to the exposure in retrospective studies. For time-varying treatment or exposure regimes, specific time points should be clearly specified and the covariates history up to and not beyond those time points should be used as population descriptors.
- CI-2 Describe population that gave rise to the effect estimate(s)**
- When conducting analyses that in some way exclude patients from the original study population, researchers should describe the final analysis population that gave rise to the effect estimate(s).
- CI-3 Precisely define the timing of the outcome assessment relative to the initiation and duration of exposure**
- To ensure that an estimate of an exposure or intervention effect corresponds to the question that researchers seek to answer, the researchers must precisely define the timing of the outcome assessment relative to the initiation and duration of the exposure.
- CI-4 Measure confounders before start of exposure and report data on confounders with study results.**
- In general, variables for use in confounding adjustment (either in the design or analysis) should be ascertained and measured prior to the first exposure to the therapy (or therapies) under study. If confounders are time varying, specific time points for the analysis of the exposure effect should be clearly specified and the confounder history up to and not beyond those time points should be used in that analysis.
- CI-5 Report the assumptions underlying the construction of propensity scores and the comparability of the resulting groups in terms of the balance of covariates and overlap**
- When conducting analyses that use propensity scores to balance covariate distributions across intervention groups, researchers should assess the overlap and balance achieved across compared groups with respect to potential confounding variables.
- CI-6 Assess the validity of the instrumental variable (i.e., how the assumptions are met) and report the balance of covariates in the groups created by the instrumental variable for all instrumental variable analyses**
- When an instrumental variable (IV) approach is used, empirical evidence should be presented describing how the variable chosen as an IV satisfies the three key properties of a valid instrument: 1) the IV influences choice of the intervention or is associated with a particular intervention because both have a common cause; 2) the IV is unrelated to patient characteristics that are associated with the outcome; and 3) the IV is not otherwise related to the outcome under study (i.e., it does not have a direct effect on the outcome apart from its effect through exposure).

9. Standards for Adaptive and Bayesian Trial Designs

Randomized trials have advantages and disadvantages when they are used to determine the comparative effectiveness of different treatments or interventions. They can provide strong evidence, but they are also often perceived as taking too long to get results or being too rigid in a rapidly changing world. Adaptive trials build upon the approaches used in most clinical trials, but they differ in that they allow changes to be made to a study while it is under way. Examples of adaptations include changing what proportion of patients are randomized to which intervention group, altering the sample size, changing the eligibility criteria, dropping or adding comparison groups, changing endpoints, and stopping early. Rather than waiting until the end of the study period to see the results and suggest changes for the next study, the changes are planned for as part of the trial design and executed based on the analyses conducted during the trial.

Many adaptive features can be implemented individually using classical statistics, often called frequentist approaches, but complex designs combining several dimensions of adaptation typically require a different statistical approach known as Bayesian analyses. Therefore, adaptive trials may also be referred to as Bayesian Trial Designs.

Recognizing the need for innovative clinical trial design, representatives from the NIH's Clinical and Translational Science Award programs have identified adaptive clinical trial design as a high-priority methodological issue "to increase the efficiency of comparative effectiveness trials" (Helfand et al. 2011). Adaptive designs are particularly appealing for PCOR because they could maintain many of the advantages of randomized clinical trials while minimizing some of the disadvantages. Adaptive methods can sometimes shorten trials. They also can increase the relevance of trial results by adjusting both the composition of patient groups and the treatments being compared. But such flexibility and efficiency have to be balanced with the risk that adaptive trials typically require a longer design period, are more complex, and are more difficult to conduct. Therefore, designing and conducting these trials require specialized expertise and experience.

Adaptive designs for trials are not new, but recently they have become more popular due, in part, to efforts to streamline drug and device development. To date, the use of adaptive trials for PCOR has been limited, with few published examples (Fiore et al. 2011; Muss et al. 2009). However, many trials

have some adaptive features—such as stopping guidelines and sample size re-estimation—that have become standard practices.

Rationale for the Standards

Adaptive trials should adhere to the principles of good design and analysis that apply to all rigorous research; however, their complexity of can make this more difficult, requiring extra attention to specific steps in the research process. Although current practice does not offer extensive guidance for adaptive trials in PCOR, the experience in therapeutics and device trials, combined with theoretical considerations, provide the basis for standards governing their design and conduct.

These studies typically require that simulations be conducted during the design phase to define the error rates, and descriptions of the design—both in protocols and published papers—must include adequate detail about the study elements and planned adaptations. Good adaptive trial design requires preplanning and specification of procedures at the outset (**AT-1**). Given the potential complexity introduced by adaptations, the timing of interim analyses and the changes that could be made based on those data should be determined before the trial starts, and the statistical approaches to evaluation and decision making should be considered and reported—and records should be maintained—so that these can be verified (**AT-2**). Similarly, standardized reporting of trials has become part of best practice and, to the extent that existing reporting guidelines (i.e., CONSORT) can be used, they should be followed and any modifications described (**AT-5**).

Other components of adaptive trials necessitate special focus: adaptation requires an infrastructure to obtain and analyze the data needed for design changes as the trial proceeds. Because this capacity is not the norm in conventional trials, it is included in the standards (**AT-4**). Adaptive trials that use Bayesian approaches require even more detailed specification of the analysis plan than is the current practice or would be required in traditional trials, both because software is not standardized and because Bayesian methods have analytic features absent in standard trials (**AT-3**).

Standards for Specific Study Designs and Methods

9. Standards for Adaptive and Bayesian Trial Designs

AT-1 Specify planned adaptations and primary analysis

The adaptive clinical trial design should be prospectively planned and the design clearly documented, including:

- All potential adaptations, including timing;
- Trial results and populations that will be used in determining each adaptation;
- Statistical models to be used; and
- Planned analysis of the primary endpoint(s).

The description of the design should be sufficiently detailed that it could be implemented from the description of procedures. The specification of the design should be completed and documented in the trial protocol before enrollment begins. This specification should include, in all but the simplest designs, a statistical analysis plan (SAP) that is separate from the trial protocol in which all necessary detail is provided regarding planned interim and final analyses. Prior specification is a prerequisite for valid and meaningful evaluation of an adaptive design.

AT-2 Evaluate statistical properties of adaptive design

While not necessary for simple designs, the statistical properties of complex adaptive clinical trial designs should be thoroughly investigated over the relevant range of important parameters or clinical scenarios (e.g., treatment effects, accrual rates, delays in the availability of outcome data, dropout rates, missing data, drift in participant characteristics over time, subgroup-treatment interactions, or violations of distributional assumptions). Statistical properties to be evaluated should include Type I error, power, and sample size distributions, as well as the precision and bias in the estimation of treatment effects. Additional performance metrics may also be evaluated (e.g., the frequency with which specific adaptations occur, the likelihood of substantial covariate imbalance, the likely adequacy of final data for subgroup and safety analyses). The programming code used to create the simulations should be retained with version control. The programming code and software used should be made available to stakeholders who have a need to know, including reviewing agencies.

AT-3 Specify structure and analysis plan for Bayesian adaptive randomized clinical trial designs

If a Bayesian adaptive design is proposed, the Bayesian structure and analysis plan for the trial must be clearly and completely specified. This should include any statistical models used either during the conduct of the trial or for the final analysis, prior probability distributions and their basis, utility functions associated with the trial's goals, and assumptions regarding exchangeability (of participants, of trials, and of other levels). Specific details should be provided as to how the prior distribution was determined and if an informative or non-informative prior was chosen. When an informative prior is used, the source of the information should be described. If the prior used during the design phase is different from the one used in the final analysis, then the rationale for this approach should be indicated. Utility functions, if employed, should be defined, and their source should be described. Computational issues, such as the choice of software, the creation and testing of custom software, and software validation, should be addressed as well. Software used for Bayesian calculations during trial design, trial execution, and final analysis must be functionally equivalent. When feasible, software or programs should be made available to relevant stakeholders for evaluation and validation.

9. Standards for Adaptive and Bayesian Trial Designs (Continued)

AT-4 Ensure clinical trial Infrastructure is adequate to support planned adaptation(s)

The clinical trial infrastructure, including centralized randomization, data collection related to the assessment and recording of key outcomes, data transmission procedures, and processes for implementing the adaptation (e.g., centralized, web-based randomization), must be able to support the planned trial. In simple adaptive trials, qualitative verification of the capabilities of the proposed trial infrastructure may be adequate. Trials with more complicated requirements, such as frequent interim analyses, require thorough testing prior to trial initiation. Such testing should involve the trial's data collection and data management procedures, the implementation of the adaptive algorithm, and methods for implementing the resulting adaptation(s). The impact on the trial's operating characteristics of delays in collecting and analyzing available outcome data should be assessed. The study plan should clarify who will perform the analyses to inform adaptation while the study is ongoing and who will have access to the results. The interim analyses should be performed and reviewed by an analytical group that is independent from the investigators who are conducting the trial. Trial investigators should remain blinded to changes in treatment allocation rates as this information provides data regarding treatment success.

AT-5 Use the CONSORT statement, with modifications, to report adaptive randomized clinical trials

The following sections of the CONSORT statement can be used to report key dimensions of adaptation:

- Adaptation of randomization probabilities (sections 8b and 13a);
- Dropping or adding study arms (sections 7b and 13a);
- Interim stopping for futility and superiority (sections 7b and 14b);
- Sample size re-estimation (sections 7a and 7b);
- Transitioning of stages (e.g., seamless Phase II/III designs) (sections 3a, 7a, 7b, and 16); and
- Modification of inclusion and exclusion criterion (sections 4a and 13a).

CONSORT sections 16, 20, and 21 may also be expanded to report additional aspects of an adaptive trial.

If the trial incorporates adaptations other than those listed above, the authors should use their judgment as to where in the CONSORT structure to include both design details and the associated results. All possible adaptations included in the prospective design, even if they did not occur, should be included in the report.

10. Standards for Studies of Diagnostic Tests

Patients, caregivers, and clinicians need specific information about the expected benefits and harms of a diagnostic test in their particular circumstances when deciding whether a test should be performed. When the research on a test is flawed, clinicians who obtain the test may under- or overestimate the likelihood that a patient has a disease. Some diagnostic tests also expose patients to unnecessary inconvenience or harm, including radiation exposure and complications from invasive procedures undertaken in response to test results. However, diagnostic testing's impact on patient outcomes has been traditionally understudied in clinical research. Studies of diagnostic tests tend not to identify all of the pertinent effects on patients, particularly long-term benefits and harms, as well as cognitive, emotional, social, and behavioral effects (Bossuyt and McCaffery 2009).

A fundamental issue in diagnostic test research is how to define the benefit of a test. Tests generate information but do not directly produce a better outcome for the patient. To improve outcomes, the test result must be used effectively—for example, by helping with a decision about which treatment or intervention to use, what lifestyle changes might avert or ameliorate disease, or what additional tests should be performed. A challenge for investigators designing a study of a diagnostic test is whether to specify the actions clinicians should take based on test results (such as observation, further testing, or treatment) or to leave those responses to the discretion of patients and their providers.

Diagnostic tests are studied through both experiments (including RCTs) and observational studies (including reviews of medical records and registries). A wide variety of observational designs has been used to assess the accuracy and impact of diagnostic tests (Lord et al. 2009). The US Food and Drug Administration and the CONSORT group offer guidance about methods for the evaluation of diagnostic tests (FDA 2007; Moher et al. 2010; Schulz et al. 2010). Other groups have recommended guidelines for reporting the results of studies of diagnostic accuracy (Bossuyt et al. 2003a,b; Whiting 2011; Whiting 2006 et al.). Standards for systematic reviews of test accuracy are also being developed (Reitsma et al. 2009; Matchar 2012; Santaguida et al. 2012; Trikalinos et al. 2012; Trikalinos and Ballion 2012; Trikalinos et al. 2012). Although these guidelines address the reporting of diagnostic or predictive accuracy studies, standards have not been established for studying the impact of diagnostic tests on subsequent care or patient outcomes.

In addition to diagnosis, medical tests have many other uses: to predict the risk of developing a disease; to establish the prognosis of a disease or condition; to predict the chance of a response or of serious adverse effects to a treatment; and, especially for imaging studies, to identify the anatomic location and extent of disease. These other uses are not covered by the current standards.

Rationale for the Standards

The diagnostic test standards (**DT-1** to **DT-5**) reflect four principles:

- Accuracy alone is often not a sufficient measure of the benefit of a test;
- Alternate tests or testing strategies should be compared in terms of effect on patient outcomes;
- The clinical context in which the test is used should be addressed, and
- The overall scientific validity and transparency of a study depends on knowing how the factors affect clinical outcomes (Ferrante di Ruffano et al. 2012) apply and supports the overall scientific validity and transparency of the study. (DT-3).

Standards for Specific Study Designs and Methods

10. Standards for Studies of Diagnostic Tests

DT-1 Specify clinical context and key elements of diagnostic test study design

A comparative evaluation of diagnostic tests should specify each of the following items and provide rationale in support of the particular choices: (a) the intended use of the test and the corresponding clinical context, including referral for additional testing, referral for additional treatments, and modification of current treatment and target populations; (b) the goal of the comparison; (c) the technical specifications of the tests as implemented in the study; (d) the approach to test interpretation; (e) the sources and process for obtaining reference standard information, when applicable; and (f) the procedures for obtaining follow-up information and determining patient outcomes, when applicable. These items ought to be specified for all designs, including observational designs (e.g., those using medical records or registries). If these items are not available directly, validated approaches to approximating these study elements from available data should be used.

DT-2 Study Design Should be Informed by Investigations of the Clinical Context of Testing

Design of comparative effectiveness studies should outline clinical pathways involving the tests and the anticipated implications of test use on downstream processes of care and patient outcomes. In the written research methods and study protocol, investigators should give examples of clinical pathways to demonstrate thorough understanding of the clinical context.

DT-3 Assess the Effect of Factors Known to Affect Diagnostic Performance and Outcomes

Studies of diagnostic tests should include an assessment of the effect of important factors known to affect test performance and outcomes, including the threshold for declaring a “positive” test result, the technical characteristics of the test and the interpreter, and the setting of care.

DT-4 Structured Reporting of Diagnostic Comparative Effectiveness Study Results

Broadly accepted checklists for reporting studies and assessing study quality, such as CONSORT, STARD, and QUADAS, should be consulted and utilized. Consult the CONSORT 2010 checklist for reporting randomized controlled trials. Consult the STARD checklist for reporting diagnostic accuracy studies. Consult the QUADAS-2 (updated in 2011) for additional guidance on reporting information that would be more useful to systematic reviews of diagnostic accuracy studies.

DT-5 Focus Studies of Diagnostic Tests on Patient-centered Outcomes, Using Rigorous Study Designs with Preference for Randomized Controlled Trials

Studies of clinical outcomes after diagnostic testing should use a prospective randomized study design when possible. If a non-randomized design is proposed, the reason for using an observational study (or modeling and simulation) should be addressed and efforts to minimize confounding documented.

11. Standards for Systematic Reviews

Systematic reviews find, assess, and synthesize results from several individual studies in order to determine what is known about the benefits and harms of specific medical interventions. Systematic reviews are used by clinicians in practice, by patients in making choices about their care, and by organizations in developing clinical practice guidelines and policies. Systematic reviews are also used to identify the gaps in the available research evidence and to outline possible topics for future research. Systematic reviews are important for PCOR because they facilitate the efficient use of existing research results and aide in targeting future work. Often, it is only by looking at a large body of evidence that it is possible to compare different health interventions (see **Research Stories: Getting off the Ventilator**).

These reviews also make it possible to determine what relevant patient questions have and have not been answered (or even asked) in research. Further, systematic reviews can serve as a vehicle for transparency, offering new insights into diseases and treatments, particularly when individual patient data are made available for pooled analyses (see **Research Stories: Aspirin for the Prevention of Colorectal Cancer**).



RESEARCH STORIES: Getting off the Ventilator

When hospital patients are put on a mechanical ventilator, it's usually a matter of life and death. But the longer people are on ventilators, the greater the likelihood they will suffer complications. Usually, hospital staff members decide when to "wean" patients from the ventilators, but some studies found that doctors underestimate the ability of patients to breathe on their own. Other studies claimed that using a protocol, a series of regimented steps, for ventilator weaning is better than staff judgment, but methodological flaws made the conclusion uncertain.

To explore this issue further, researchers performed a systematic review of 11 studies (including almost 2,000 patients) that compared weaning that uses or doesn't use protocols for reducing the duration of mechanical ventilation in critically ill adult patients. The analysis (Blackwood et al. 2011) indicated that a weaning protocol, as opposed to staff judgment, reduced the average time on the ventilator by 20 to 36 hours and time in the intensive care unit by about a day. In most cases, weaning protocols were better than staff judgments.



RESEARCH STORIES: Aspirin for the Prevention of Colorectal Cancer

Since the 1990s, observational studies, such as cohort studies, have shown that patients who regularly use aspirin suffer a lower-than-average risk of colorectal cancer. Because the protective benefit takes more than 10 years to appear, even long-term randomized controlled trials like the Physicians' Health Study could not replicate these findings. To address the limitation of existing trial data, investigators conducted a systematic review of four randomized trials of daily aspirin versus placebo that had originally been designed to evaluate the benefits of aspirin for preventing heart attacks and strokes. The investigators took their meta-analysis a step further by obtaining the original patient data from those trials and using national cancer registries in the United Kingdom or Sweden to follow patients for up to 20 years after they started taking aspirin or a placebo.

The investigators found that daily aspirin reduced the 20-year risk of colorectal cancer by 24 percent and colorectal cancer mortality by 35 percent (Rothwell et al. 2011, 2012). Patients did not necessarily continue taking daily aspirin after the original randomized controlled trials finished; an average of six years of daily aspirin during the trials was sufficient to reduce the rate of colorectal cancer and its mortality. Among patients who were assigned randomly to take aspirin for at least five years, higher dose aspirin failed to improve on the benefit of a relatively low dose (75 mg to 300 mg per day).

By linking trial data with national cancer registries, the investigators were able to answer a research question more efficiently; a new randomized trial to address the question would have required 20 years and also millions of dollars in additional funding.

Many organizations and individuals conduct systematic reviews. However, the processes used to conduct these reviews and their overall quality can vary. The search for evidence may be more or less exhaustive, how the included studies are evaluated may differ, and there may be errors when data are collected and combined from different studies.

Rationale for the Standards

In 2011, the IOM released a report titled *Finding What Works in Health Care: Standards for Systematic Reviews* (IOM 2011). PCORI has concluded that these standards are largely acceptable. The included standards were developed by a credible panel based on a broad review that considered and incorporated existing authoritative sources (e.g., Cochrane, AHRQ EPC program).

The IOM standards are designed to contribute to explicit methods, consistent application, and the opportunity for public review so that users can link judgments, decisions, or actions to the data on

which they are based. Additionally, they are intended to increase objectivity, minimize bias, improve reproducibility, and lead to more complete reporting. The IOM standards are appropriate for inclusion in PCORI standards because they aim to ensure patient-centeredness in every aspect of conducting systematic reviews of clinical effectiveness (**SR-1**).

There is a lack of empirical evidence to support many common practices in conducting systematic reviews, even some that are considered best practices. Analyses to assess the consequences and value of the IOM standards have not been conducted. Because high-quality systematic reviews can be produced by teams that do not completely conform to all IOM standards, the need for some of these standards may vary across topics and situations.

Standards for Specific Study Designs and Methods

11: Standards for Systematic Reviews

SR-1 Adopt the Institute of Medicine (IOM) standards for systematic reviews of comparative effectiveness research, with some qualifications

Systematic reviews are used to answer questions based on comprehensive consideration of all the pertinent evidence, and can also identify the gaps in evidence and how they might be resolved. Standards for systematic reviews are currently in use, but credible authorities, such as the Cochrane Collaboration and the Agency for Healthcare Research and Quality (AHRQ), vary somewhat in their recommended standards. The IOM recently issued standards that draw broadly from available sources. PCORI endorses these standards but recognizes that there can be flexibility in the application of some standards without compromising the validity of the review, specifically:

- Searches for studies reported in languages other than English are not routinely recommended, but may be appropriate to some topics;
- Dual screening and date abstraction are desirable, but fact-checking may be sufficient. Quality control procedures are more important than dual review per se; and
- Independent librarian peer review of the search strategy is not required; internal review by experienced researchers is sufficient.

Institute of Medicine (IOM). 2011. *Finding What Works in Health Care: Standards for Systematic Reviews*. Washington, DC: The National Academies Press.

This page intentionally left blank.

SECTION IV: THE CONTEXT FOR IMPLEMENTING THE METHODOLOGY STANDARDS AND NEXT STEPS

The Context for Implementing the Methodology Standards

Good research practices are a required foundation for patient-centered outcomes research (PCOR) methodology standards. One of the most important components of good practices is a commitment to transparency, which enables other researchers to verify findings. Many of the PCORI Methodology Standards promote transparency by requiring detailed protocols when researchers propose research and compliance with guidelines when they register study participants and report results. Not only can these requirements help PCORI and others judge the quality and relevance of proposed research plans, but they may also help protect against practices—such as selective reporting—that can distort or misrepresent research results. PCORI requires, in addition, that researchers assess appropriate methods for data sharing and give proper credit to those sharing protocols, code, and data.

The value of systematic reviews, which often inform future research needs, depend on the degree to which evidence is reported fully and in an unbiased manner. Credible standards for conducting systematic reviews specific to clinical effectiveness recognize that “reporting biases, particularly publication bias and selective reporting of trial outcomes and analyses, present the greatest obstacle to obtaining a complete collection of relevant information on the effectiveness of health care interventions”(IOM 2011). A significant next step for PCORI is to promote policies that can remove or overcome this obstacle, not only in the research the institute supports but throughout the broader clinical research community.

Next Steps

Developing Standards

This first set of methodology standards for PCOR is an important milestone, but not the destination. The legislation establishing PCORI directs that these standards be periodically updated. PCORI expects that the scope of the standards will be widened to include the full spectrum of PCOR questions and approaches. This task comprises: 1) refining the methods used to develop the standards, including improvements to the methods used to identify, evaluate, and synthesize existing standards; 2) refining the methods for developing new standards in areas where there are currently no standards; and 3)

reviewing the empirical evidence supporting existing and proposed standards and evaluating their usefulness in specifying appropriate research methods. As a core function of the PCORI Methodology Committee, further development of the standards is a prominent part of the blueprint for future work. Below is a partial listing of specific actions that PCORI intends to take. Further details about this agenda are included in **Appendix C: Recommended Actions and Research Recommendations**.

- Expand the inventory of research methods relevant to PCOR for which standards are needed;
- Distinguish between standards that are minimum requirements and those that may be desirable or best practice but not required;
- Specify and support new research to strengthen methods relevant to CER and PCOR; and
- Refine processes to use members of the PCORI Methodology Committee, PCORI scientific staff, external groups (e.g., Institute of Medicine, AHRQ, and professional societies), consultants, and other stakeholders in locating and assessing standards prepared by other groups and developing new standards.

Evidence in a number of areas relevant to the standards is limited, and the standards will evolve as additional information becomes available. There are three important gaps in knowledge related to patient engagement:

- What are the consequences of patient engagement in research on health decisions and clinical outcomes?
- What are the specific consequences of patient engagement on the research process?
- Which patient engagement methods are most effective, and for which populations?

PCORI is interested in advancing the science of patient-centered study design and patient and stakeholder engagement, dissemination, and implementation. Particular areas of interest include understanding optimal approaches to engaging patients and other stakeholders throughout the research continuum; understanding how such engagement affects study design and outcomes; improving strategies for recruiting and retaining patients and other stakeholders, especially those who are historically underrepresented or hard to reach; and refining approaches to minimize missing patient-reported data.

Supporting Adoption of Methodology Standards

PCORI is pursuing a comprehensive, coordinated approach to promote the wide use of its methodology standards. It is engaging a broad range of stakeholders who might use the standards; collaborating with other organizations and initiatives to strengthen research practices and facilitate use of the standards; and creating reporting and surveillance mechanisms. Future activities might include the development of training resources, checklists, and other tools to support researchers' decision making and practices, as well as checklists and other decision-support tools for peer reviewers. Other initiatives will include outreach to research, clinician, professional, and public audiences to promote use and adoption of best practices for PCOR.

This page left intentionally blank.



Appendices

Revised Draft for Board Acceptance

November 2013

APPENDIX A: PCORI METHODOLOGY STANDARDS

Cross-Cutting Standards for PCOR

1: Standards for Formulating Research Questions

RQ-1 Identify gaps in evidence

Gap analysis and systematic reviews should be used to support the need for a proposed study. If a systematic review is not available, a systematic review should be performed using accepted standards in the field (see standard SR-1), or a strong rationale should be presented for proceeding without a systematic review. In the case where a systematic review is not possible, the methods used to review the literature should be explained and justified.

RQ-2 Develop a formal study protocol

Studies should include a formal protocol specifying at least one purpose for which the data were collected (e.g., effectiveness, safety, natural history of disease, quality improvement); data sources and linkage plans, if any; data feasibility and quality, measure(s) of effect; and use of any standardized data dictionaries (nationally or internationally accepted).

RQ-3 Identify specific populations and health decision(s) affected by the research

To produce information that is meaningful and useful to people when making specific health decisions, research proposals and protocols should describe: 1) the specific health decision the research is intended to inform; 2) the specific population for whom the health decision is pertinent; and 3) how study results will inform the health decision.

RQ-4 Identify and assess participant subgroups

In designing studies, researchers should identify participant subgroups of interest and, where feasible, design the study with adequate precision and power to reach conclusions specific to these subgroups. In addition, subgroup information should be reported for later systematic reviews.

RQ-5 Select appropriate interventions and comparators

When evaluating an intervention, the comparator treatment(s) must be chosen to enable accurate evaluation of effectiveness or safety compared to other viable options for similar patients. Researchers should make explicit what the comparators are and how they were selected, focusing on clearly describing how the chosen comparator(s) define the causal question, reduce the potential for biases, and allow direct comparisons. Generally, non-use (or no specific treatment) comparator groups should be avoided unless no specific treatment is a likely option in standard care.

RQ-6 Measure outcomes that people representing the population of interest notice and care about

Identify and include outcomes the population of interest notices and cares about (e.g., survival, function, symptoms, health-related quality of life) and that inform an identified health decision. Define outcomes

clearly, especially for complex conditions or outcomes that may not have established clinical criteria. Provide information that supports the selection of outcomes as meeting the criteria of “patient-centered” and “relevant to decision makers,” such as patient and decision-maker input from meetings, surveys, or published studies. Select outcomes based on input directly elicited from patient informants and people representative of the population of interest, either in previous studies or in the proposed research.

2: Standards Associated with Patient-Centeredness

PC-1 Engage people representing the population of interest and other relevant stakeholders in ways that are appropriate and necessary in a given research context

People representing the population of interest include individuals who have the condition or who are at risk of the condition and, as relevant, their surrogates or caregivers. Other relevant stakeholders may include clinicians, administrators, policy makers, or others involved in healthcare decision making.

Stakeholders can be engaged in the processes of:

- Formulating research questions;
- Defining essential characteristics of study participants, comparators, and outcomes;
- Identifying and selecting outcomes that the population of interest notices and cares about (e.g., survival, function, symptoms, health-related quality of life) and that inform decision making relevant to the research topic;
- Monitoring study conduct and progress; and
- Designing/suggesting plans for dissemination and implementation activities.

When applicable, research proposals should describe how these stakeholders will be identified, recruited, and retained. If engagement is not necessary or appropriate in these processes, explain why.

PC-2 Identify, select, recruit, and retain study participants representative of the spectrum of the population of interest and ensure that data are collected thoroughly and systematically from all study participants

Research proposals and subsequent study reports should describe: 1) the plan to ensure representativeness of participants; 2) how participants are identified, selected, recruited, enrolled, and retained in the study to reduce or address the potential impact of selection bias; 3) efforts employed to maximize adherence to agreed-on enrollment practices; and 4) methods used to ensure unbiased and systematic data collection from all participants.

If the population of interest includes people who are more difficult to identify, recruit, and/or retain than other study populations (for example, individuals historically underrepresented in healthcare research such as those with multiple disease conditions, low literacy, low socioeconomic status, or poor healthcare access, as well as racial and ethnic minority groups and people living in rural areas), then specify plans to address population-unique issues for participant identification, recruitment, and retention.

PC-3 Use patient-reported outcomes when patients or people at risk of a condition are the best source of information

When patients or people at risk of a condition are the best source of information regarding outcomes of interest, then the study should employ patient-reported outcome (PRO) measures in lieu of, or in addition to, measures derived from other sources. Proposals should describe: 1) the concept(s) underlying each PRO measure (e.g., symptom or impairment) and how it is meaningful to, and noticed by, patients in the population of interest; 2) how the concept relates to the health decisions the study is designed to inform; 3) how the PRO measure was developed, including how patients were involved in the development; and 4) evidence of measurement properties including content validity, construct validity, reliability, responsiveness to change over time, and score interpretability, including meaningfulness of score changes in the population of interest with consideration of important subgroups. If these measurement properties are not known, a plan for establishing the properties must be provided. Caregiver reports may be appropriate if the patient cannot self-report the outcomes of interest. If PROs are not planned for use in the study, justification must be provided.

PC-4 Support dissemination and implementation of study results

Support dissemination and implementation of study results by suggesting strategies, indicating clinical and policy implications, and working with patients or organizations to report results in a manner understandable to each target audience.

3: Standards for Data Integrity and Rigorous Analyses

IR-1 Assess data source adequacy

In selecting variables for confounding adjustment, researchers should assess the suitability of the data source in terms of its ability to assure robust capture of needed covariates.

IR-2 Describe data linkage plans, if applicable

For studies involving linkage of patient data from two or more sources (including registries, data networks, and others), describe 1) each data source and its appropriateness, value, and limitations for addressing specific research aims; 2) any additional requirements that may influence successful linkage, such as information needed to match patients, selection of data elements, and definitions used; and 3) the procedures and algorithm(s) employed in matching patients, including the success, limitations, and any validation of the matching algorithm.

IR-3 A priori, specify plans for data analysis that correspond to major aims

Researchers should describe the analytic approaches that will be used to address the major research aims prior to data collection. These include definitions of key exposures, endpoints, and covariates. Also identify patient subgroups of interest, plans (if any) for how new subgroups of interest will be identified or how analysis plans may be adapted based on changing needs and scientific advances, and plans for how missing data will be handled.

IR-4 Document validated scales and tests

Studies should include documentation of the name of the scales and tests selected, reference(s), characteristics of the scale, and psychometric properties.

IR-5 Use sensitivity analyses to determine the impact of key assumptions

The results of these sensitivity analyses should be reflected in the interpretation of results.

IR-6 Provide sufficient information in reports to allow for assessments of the study's internal and external validity

Reporting guidelines for specific designs can be found at the [EQUATOR Network website \(www.equator-network.org\)](http://www.equator-network.org). This website has brought together all reporting guidelines that have been developed using formal approaches, many of which have been adopted by journals, such as CONSORT (for randomized clinical trials), STARD (for diagnostic tests), and STROBE (for observational studies).

4: Standards for Preventing and Handling Missing Data

MD-1 Describe methods to prevent and monitor missing data

Investigators should explicitly anticipate potential problems of missing data. The study protocol should contain a section that addresses missing data issues and steps taken in study design and conduct to monitor and limit the impact of missing data. Missingness can occur from patient dropout, failure to provide data, and/or administrative or data management issues. As relevant, the protocol should include the anticipated amount of and reasons for missing data, as well as plans to follow up with participants. This standard applies to all study designs for any type of research question.

MD-2 Describe statistical methods to handle missing data

Statistical methods for handling missing data should be pre-specified in study protocols. The reasons for missing data should be considered in the analysis. The plausibility of the assumptions associated with the approach should be assessed. A discussion of the potential ramifications of the approach to missing data on the results should be provided. This standard applies to all study designs for any type of research question.

MD-3 Use validated methods to deal with missing data that properly account for statistical uncertainty due to missingness

Statistical inference of intervention effects or measures of association should account for statistical uncertainty attributable to missing data. This means that methods used for imputing missing data should have valid Type I error rates and that confidence intervals should have the nominal coverage properties. This standard applies to all study designs for any type of research question. Bayesian methods and methods such as multiple imputation satisfy this condition, along with various likelihood-based and other validated methods. Single imputation methods like last observation carried forward and baseline observation carried forward are discouraged as the primary approach for handling missing data in the analysis. If investigators do use single-based imputation methods, they must provide a compelling scientific rationale as to why the method is appropriate.

MD-4 Record and report all reasons for dropout and missing data, and account for all patients in reports

Whenever a participant drops out of a research study, the investigator should document the following: 1) the specific reason for dropout, in as much detail as possible; 2) who decided that the participant would drop out; and 3) whether the dropout involves some or all types of participation. Investigators should attempt to continue to collect information on key outcomes for participants unless consent is withdrawn. This standard applies to all prospective study designs that aim to assess intervention effectiveness. All participants included in the study should be accounted for in the report, whether or not they are included in the analysis. Describe and justify any planned reasons for excluding participants from analysis.

MD-5 Examine sensitivity of inferences to missing data methods and assumptions, and incorporate into interpretation

Examining sensitivity to the assumptions about the missing data mechanism (i.e., sensitivity analysis) should be a mandatory component of the study protocol, analysis, and reporting. This standard applies to all study designs for any type of research question. Statistical summaries should be used to describe missing data in studies, including a comparison of baseline characteristics of units (e.g., patients, questions, or clinics) with and without missing data. These quantitative results should be incorporated into the interpretation of the study and reflected in the discussion section and possibly the abstract.

5: Standards for Heterogeneity of Treatment Effects

HT-1 State the goals of HTE analyses

State the inferential goal of each HTE analysis, specifying how it is related to the topic of the research, translate this into an analytic approach, and highlight the linkages between the two. Identify analyses as hypothesis driven (sometimes denoted confirmatory), or hypothesis generating (sometime denoted exploratory).

HT-2 For all HTE analyses, pre-specify the analysis plan; for hypothesis-driven HTE analyses, pre-specify hypotheses and supporting evidence base

The study protocol should unambiguously pre-specify planned HTE analyses. Pre-specification of hypothesis-driven HTE analyses should include a clear statement of the hypotheses the study will evaluate, including how groups will be defined (e.g., by multivariate score or stratification) and outcome measures, and the direction of the expected treatment effects. The pre-specified hypotheses should be based on prior evidence, which should be described clearly in the study protocol and published paper.

HT-3 All HTE claims must be based on appropriate statistical contrasts among groups being compared, such as interaction tests or estimates of differences in treatment effect

A common error in HTE analyses is to claim differences in treatment effect when one group shows a statistically significant treatment effect and another does not. To claim differences in treatment effect among subgroups, appropriate statistical methods must be used to directly contrast them. Such contrasts include, but are not limited to, interaction tests, differences in treatment effect estimates with standard errors, or a variety of approaches to adjusting the estimated subgroup effect, such as

Bayesian shrinkage estimates. Within each subgroup level, studies should present the treatment effect estimates and measures of variability.

HT-4 For any HTE analysis, report all pre-specified analyses and, at minimum, the number of post hoc analyses, including all subgroups and outcomes analyzed

Protocols and study reports must report the exact procedures used to explore HTE, including data mining or any automatic regression approaches. HTE analyses should clearly report the procedures by which subgroups were defined (e.g., by categorical predictors or continuous risk scores) and the effective number of subgroups and outcomes examined. If a non-prespecified stratum or subgroup is claimed to show a treatment effect that is different from others, methods should be used that account for the number of contrasts examined. These methods include, but are not limited to, p-value adjustment, false discovery rates, Bayesian shrinkage estimates, adjusted confidence intervals, and validation methods (internal or external).

Standards for Specific Study Designs and Methods

6: Standards for Data Registries

DR-1 Requirements for the design and features of registries

Registries established for conducting PCOR must have the following characteristics to facilitate the collection and aggregation of usable data, to ensure appropriate privacy and confidentiality, to document changes to the registry protocol, and to guide robust analyses that include important confounders.

A. Patient Follow-up

The objective(s) of the registry should determine the type, extent, and length of patient follow-up. Describe what triggers the follow-up, the follow-up measures, and the last contact with the patient. Ensure that the planned follow-up time is adequate to address the main objective and that planned patient-retention efforts are suitable to the target population and anticipated challenges. Describe expected loss to follow-up and potential effect on the results, including possible biases resulting from differential loss.

B. Data Safety and Security

Registry custodians should provide transparency for institutional review boards by describing data use agreements, informed consent, data security, and approaches to protecting security including risk of re-identification of patients. If using previously collected data, describe how these address the risk of re-identification of patients and the actual use of data compared with the originally designed and consented use of the data.

C. Data Quality Assurance

A quality assurance plan for registries should address: 1) structured training tools for data abstractors; 2) use of data quality checks for ranges and logical consistency for key exposure and outcome variables and covariates; and 3) data review and verification procedures, including source data verification plans and validation statistics focused on the key exposure and outcome variables and covariates for which sites may be especially challenged. A risk-based approach to quality assurance is advisable, focused on variables of greatest importance.

D. Document and Explain Any Modifications to the Protocol

Modifications to a registry protocol may be necessary for a variety of reasons. When modifications are necessary, they should be explained, documented, and made available to anyone planning to use the registry data.

E. Consistent Data Collection

Clear, operational definitions of data elements should be provided. Create and distribute standard instructions to data collectors. Use standardized data element definitions and/or data dictionaries whenever possible. When creating a new registry, published literature should be reviewed to identify existing, widely used definitions before drafting new definitions.

F. Systematic Patient Enrollment and Follow-up

Enroll patients systematically and follow them in as unbiased a manner as possible, using similar procedures at all participating sites. Describe how patients and providers were recruited into the study to allow the impact of selection bias to be clearly understood; for example, by explaining whether the sampling was population-based or otherwise and any efforts employed to confirm the quality of adherence to agreed-on enrollment practices.

G. Monitor and Minimize Loss to Follow-up

Monitor loss to follow-up to ensure that follow-up is reasonably complete for the main objective. Minimizing loss to follow-up requires having a target and advance planning for what actions will be employed in the event that this target is in jeopardy. At the outset of the registry, develop a patient retention plan that documents when a patient will be considered lost to follow-up and what actions will be taken to minimize such loss. At the enrollment visit, consider collecting multiple types of contact information (e.g., telephone, mailing address, and email address) for the patient, as well as collecting contact information for an alternate contact if the patient cannot be reached directly. Verify contact information at each subsequent visit and update as needed. When a patient misses a visit, contact the patient following a standard protocol (e.g., phone call one day after missed visit, email one week after missed visit). If the patient withdraws from the registry, attempt to document the reason for withdrawal so that issues can be identified and addressed (e.g., overly burdensome patient-reported outcome measures). Efforts at minimizing loss to follow-up should be tempered by considerations and sensitivity to repeated intrusions on patients and to the health conditions and interventions under study. Consider collecting enough information to permit accurate linkage with other data sources, such as the National Death Index, for long-term follow-up.

H. Collect Data to Address Confounding

Registries should identify important potential confounders during the planning phase and collect reasonably sufficient data on these potential confounders to facilitate the use of appropriate statistical techniques during the analysis phase.

DR-2 Selection and use of registries

Researchers planning PCOR studies relying on registries must ensure that these meet the requirements contained in Standard DR-1 and must document each required feature of the registry(s) to be used (e.g., in an appendix to the funding application or study protocol). Deviations from the requirements should be justified by explaining why a required feature is not feasible or not necessary to achieve the overall goals of Standard DR-1.

DR-3 Robust analysis of confounding factors

In studies that use registries to evaluate the comparative effectiveness or safety of interventions, investigators should select an approach for adjusting for known and measured confounders, such as multivariable regression analysis or propensity scores to create matched comparison groups, or an instrumental variable analysis if a valid instrument is available. It is also desirable to examine the robustness of the results through sensitivity analyses focused on testing key assumptions and evaluating the likely impact of unmeasured confounders. The rationale for using selected techniques, any assumptions made, and the strengths and limitations of the techniques should be described in reports of the study findings to allow for informed interpretation of the results.

7. Standards for Data Networks as Research-Facilitating Structures

DN-1 Requirements for the design and features of data networks

Data networks established for conducting PCOR must have the following characteristics to facilitate valid, useable data and to ensure appropriate privacy, confidentiality, and intellectual property protections:

- A. **Data Integration Strategy**—In order for equivalent data elements from different sources to be harmonized (treated as equivalent), processes should be created and documented that either 1) transform and standardize data elements prior to analysis or 2) make transformation logic available that can be executed when data are extracted. The selected approach should be based on an understanding of the research domain of interest.
- B. **Risk Assessment Strategy**—Data custodians should measure the risk of re-identification of data and apply algorithms to ensure that the desired level of confidentiality is achieved to meet the need of the particular PCOR application.
- C. **Identity Management and Authentication of Individual Researchers**—Develop reliable processes for verifying credentials of researchers who are granted access to a distributed research network and for authenticating them.
- D. **Intellectual Property Policies**—A research network should develop policies for the handling and dissemination of intellectual property (IP); networks should also have an ongoing process for reviewing and refreshing those policies. IP can include data, research databases, papers, reports, patents, and/or products resulting from research using the network. Guidelines should balance 1) minimizing impediments to innovation in research processes and 2) making the results of research widely accessible, particularly to the people who need them the most.
- E. **Standardized Terminology Encoding of Data Content**—The data contents should be represented with standardized terminology systems to ensure that their meaning is unambiguously and consistently understood by parties using the data.
- F. **Metadata Annotation of Data Content**—Semantic and administrative aspects of data contents should be annotated with a set of metadata items. Metadata annotation helps to correctly identify the intended meaning of a data element and facilitates an automated compatibility check among data elements.
- G. **Common Data Model**—Individual data items should be assembled into a contextual environment that shows close or distant association among data. A common data model (CDM) specifies necessary data items that need to be collected and shared across participating institutes, clearly represents these associations and relationships among data elements, and promotes correct interpretation of the data content.

DN-2 Selection and use of data networks

Researchers planning PCOR studies relying on data networks must ensure that these networks meet the requirements contained in Standard DN-1, and they must document each required feature of the data network(s) to be used (e.g., in an appendix to the funding application or study protocol). Deviations from the requirements should be justified by explaining why a required feature is not feasible or not necessary to achieve the overall goals of standard DN-1.

8. Standards for Causal Inference Methods**CI-1 Define analysis population using covariate histories**

Decisions about whether patients are included in an analysis should be based on information available at each patient's time of study entry in prospective studies or on information from a defined time period prior to the exposure in retrospective studies. For time-varying treatment or exposure regimes, specific time points should be clearly specified and the covariates history up to and not beyond those time points should be used as population descriptors.

CI-2 Describe population that gave rise to the effect estimate(s)

When conducting analyses that in some way exclude patients from the original study population, researchers should describe the final analysis population that gave rise to the effect estimate(s).

CI-3 Precisely define the timing of the outcome assessment relative to the initiation and duration of exposure

To ensure that an estimate of an exposure or intervention effect corresponds to the question that researchers seek to answer, the researchers must precisely define the timing of the outcome assessment relative to the initiation and duration of the exposure.

CI-4 Measure confounders before start of exposure and report data on confounders with study results.

In general, variables for use in confounding adjustment (either in the design or analysis) should be ascertained and measured prior to the first exposure to the therapy (or therapies) under study. If confounders are time varying, specific time points for the analysis of the exposure effect should be clearly specified and the confounder history up to and not beyond those time points should be used in that analysis.

CI-5 Report the assumptions underlying the construction of propensity scores and the comparability of the resulting groups in terms of the balance of covariates and overlap

When conducting analyses that use propensity scores to balance covariate distributions across intervention groups, researchers should assess the overlap and balance achieved across compared groups with respect to potential confounding variables.

CI-6 Assess the validity of the instrumental variable (i.e., how the assumptions are met) and report the balance of covariates in the groups created by the instrumental variable for all instrumental variable analyses

When an instrumental variable (IV) approach is used, empirical evidence should be presented describing how the variable chosen as an IV satisfies the three key properties of a valid instrument: 1) the IV influences choice of the intervention or is associated with a particular intervention because both have a common cause; 2) the IV is unrelated to patient characteristics that are associated with the outcome; and 3) the IV is not otherwise related to the outcome under study (i.e., it does not have a direct effect on the outcome apart from its effect through exposure).

9. Standards for Adaptive and Bayesian Trial Designs

AT-1 Specify planned adaptations and primary analysis

The adaptive clinical trial design should be prospectively planned and the design clearly documented, including:

- All potential adaptations, including timing;
- Trial results and populations that will be used in determining each adaptation;
- Statistical models to be used; and
- Planned analysis of the primary endpoint(s).

The description of the design should be sufficiently detailed that it could be implemented from the description of procedures. The specification of the design should be completed and documented in the trial protocol before enrollment begins. This specification should include, in all but the simplest designs, a statistical analysis plan (SAP) that is separate from the trial protocol in which all necessary detail is provided regarding planned interim and final analyses. Prior specification is a prerequisite for valid and meaningful evaluation of an adaptive design.

AT-2 Evaluate statistical properties of adaptive design

While not necessary for simple designs, the statistical properties of complex adaptive clinical trial designs should be thoroughly investigated over the relevant range of important parameters or clinical scenarios (e.g., treatment effects, accrual rates, delays in the availability of outcome data, dropout rates, missing data, drift in participant characteristics over time, subgroup-treatment interactions, or violations of distributional assumptions). Statistical properties to be evaluated should include Type I error, power, and sample size distributions, as well as the precision and bias in the estimation of treatment effects.

Additional performance metrics may also be evaluated (e.g., the frequency with which specific adaptations occur, the likelihood of substantial covariate imbalance, the likely adequacy of final data for subgroup and safety analyses). The programming code used to create the simulations should be retained with version control. The programming code and software used should be made available to stakeholders who have a need to know, including reviewing agencies.

AT-3 Specify structure and analysis plan for Bayesian adaptive randomized clinical trial designs

If a Bayesian adaptive design is proposed, the Bayesian structure and analysis plan for the trial must be clearly and completely specified. This should include any statistical models used either during the conduct of the trial or for the final analysis, prior probability distributions and their basis, utility functions associated with the trial's goals, and assumptions regarding exchangeability (of participants, of trials, and of other levels). Specific details should be provided as to how the prior distribution was determined and if an informative or non-informative prior was chosen. When an informative prior is used, the source of the information should be described. If the prior used during the design phase is different from the one used

in the final analysis, then the rationale for this approach should be indicated. Utility functions, if employed, should be defined, and their source should be described. Computational issues, such as the choice of software, the creation and testing of custom software, and software validation, should be addressed as well. Software used for Bayesian calculations during trial design, trial execution, and final analysis must be functionally equivalent. When feasible, software or programs should be made available to relevant stakeholders for evaluation and validation.

AT-4 Ensure clinical trial Infrastructure is adequate to support planned adaptation(s)

The clinical trial infrastructure, including centralized randomization, data collection related to the assessment and recording of key outcomes, data transmission procedures, and processes for implementing the adaptation (e.g., centralized, web-based randomization), must be able to support the planned trial. In simple adaptive trials, qualitative verification of the capabilities of the proposed trial infrastructure may be adequate. Trials with more complicated requirements, such as frequent interim analyses, require thorough testing prior to trial initiation. Such testing should involve the trial's data collection and data management procedures, the implementation of the adaptive algorithm, and methods for implementing the resulting adaptation(s). The impact on the trial's operating characteristics of delays in collecting and analyzing available outcome data should be assessed. The study plan should clarify who will perform the analyses to inform adaptation while the study is ongoing and who will have access to the results. The interim analyses should be performed and reviewed by an analytical group that is independent from the investigators who are conducting the trial. Trial investigators should remain blinded to changes in treatment allocation rates as this information provides data regarding treatment success.

AT-5 Use the CONSORT statement, with modifications, to report adaptive randomized clinical trials

The following sections of the CONSORT statement can be used to report key dimensions of adaptation:

- Adaptation of randomization probabilities (sections 8b and 13a);
- Dropping or adding study arms (sections 7b and 13a);
- Interim stopping for futility and superiority (sections 7b and 14b);
- Sample size re-estimation (sections 7a and 7b);
- Transitioning of stages (e.g., seamless Phase II/III designs) (sections 3a, 7a, 7b, and 16); and
- Modification of inclusion and exclusion criterion (sections 4a and 13a).

CONSORT sections 16, 20, and 21 may also be expanded to report additional aspects of an adaptive trial.

If the trial incorporates adaptations other than those listed above, the authors should use their judgment as to where in the CONSORT structure to include both design details and the associated results. All possible adaptations included in the prospective design, even if they did not occur, should be included in the report.

10. Standards for Studies of Diagnostic Tests

DT-1 Specify clinical context and key elements of diagnostic test study design

A comparative evaluation of diagnostic tests should specify each of the following items and provide rationale in support of the particular choices: 1) the intended use of the test and the corresponding clinical context, including referral for additional testing, referral for additional treatments, and modification of current treatment and target populations; 2) the goal of the comparison; 3) the technical

specifications of the tests as implemented in the study; 4) the approach to test interpretation; 5) the sources and process for obtaining reference standard information, when applicable; and 6) the procedures for obtaining follow-up information and determining patient outcomes, when applicable. These items ought to be specified for all designs, including observational designs (e.g., those using medical records or registries). If these items are not available directly, validated approaches to approximating these study elements from available data should be used.

DT-2 Study design should be informed by investigations of the clinical context of testing

Design of comparative effectiveness studies should outline clinical pathways involving the tests and the anticipated implications of test use on downstream processes of care and patient outcomes. In the written research methods and study protocol, investigators should give examples of clinical pathways to demonstrate thorough understanding of the clinical context.

DT-3 Assess the effect of factors known to affect diagnostic performance and outcomes

Studies of diagnostic tests should include an assessment of the effect of important factors known to affect test performance and outcomes, including the threshold for declaring a “positive” test result, the technical characteristics of the test and the interpreter, and the setting of care.

DT-4 Structured reporting of diagnostic comparative effectiveness study results

Broadly accepted checklists for reporting studies and assessing study quality, such as CONSORT, STARD, and QUADAS, should be consulted and utilized. Consult the CONSORT 2010 checklist for reporting randomized controlled trials. Consult the STARD checklist for reporting diagnostic accuracy studies. Consult the QUADAS-2 (updated in 2011) for additional guidance on reporting information that would be more useful to systematic reviews of diagnostic accuracy studies.

DT-5 Focus studies of diagnostic tests on patient-centered outcomes, using rigorous study designs with preference for randomized controlled trials

Studies of clinical outcomes after diagnostic testing should use a prospective randomized study design when possible. If a non-randomized design is proposed, the reason for using an observational study (or modeling and simulation) should be addressed and efforts to minimize confounding documented.

11: Standards for Systematic Reviews

SR-1 Adopt the Institute of Medicine (IOM) standards for systematic reviews of comparative effectiveness research, with some qualifications

Systematic reviews are used to answer questions based on comprehensive consideration of all the pertinent evidence, and can also identify the gaps in evidence and how they might be resolved. Standards for systematic reviews are currently in use, but credible authorities, such as Cochrane and the Agency for Healthcare Research and Quality (AHRQ), vary somewhat in their recommended standards. The IOM recently issued standards that draw broadly from available sources. The PCORI Methodology Committee endorses these standards but recognizes that there can be flexibility in the application of some standards without compromising the validity of the review, specifically:

- Searches for studies reported in languages other than English are not routinely recommended, but may be appropriate to some topics;

- Dual screening and date abstraction are desirable, but fact-checking may be sufficient. Quality control procedures are more important than dual review per se; and
- Independent librarian peer review of the search strategy is not required; internal review by experienced researchers is sufficient.

IOM (Institute of Medicine). *Finding What Works in Health Care: Standards for Systematic Reviews*. Washington, DC: The National Academies Press, 2011.

APPENDIX B: RESPONSE TO PUBLIC COMMENT

Background

In an effort to promote transparency, meet legislative mandates, and increase the usefulness of its initial attempt to create methodology standards, PCORI solicited input from stakeholders on the draft PCORI Methodology Report (including the PCORI Methodology Standards) through a public comment period that ran from July to September 2012. There were 124 unique submissions and 1,487 unique comments from a wide variety of stakeholder groups. Individuals who identified themselves as researchers or industry representatives provided more than half of all comments (28 and 27 percent, respectively, for a total of 55 percent). Nine percent of the comments were submitted by patients, caregivers, or patient advocates. Other submissions came from organizations, professional associations, advocacy organizations, academic institutions, clinicians, and policy makers. PCORI thanks the individuals, organizations, and institutions that took time to provide the many thoughtful and meaningful suggestions.

These comments have been extremely useful to PCORI in revising the report and methodology standards, as well as guiding PCORI's plans for improving the science of comparative effectiveness research.

After the close of the public comment period, PCORI revised many of the methodology standards, and reduced the total number of standards, based on consideration of those comments that specifically addressed the standards themselves. The revised standards were adopted by the PCORI Board of Governors and published in December 2012. However, a revised version of the full methodology report was not released at that time.

Subsequently, the PCORI Methodology Committee and staff considered the public comments that addressed the report and made extensive revisions, fully addressing those comments. This appendix summarizes all of the changes to both the PCORI Methodology Standards in 2012 and the PCORI Methodology Report in 2013. In the process of revising the report based on public comment, minor editorial changes were made to improve the clarity, but not change the meaning, of the standards. The PCORI Methodology Committee oversaw and accepted the minor revisions.

Overview of Public Comments

The 1,487 individual comments were assembled into a combined document that did not include commenter identity or affiliation. The initial coding and analysis of comments was completed by an external contractor, who also aggregated the comments by theme and category. Each comment was considered and its recommendation was considered in the context of PCORI's broader methodology agenda.

The comments fell into three broad categories: 1) suggestions for PCORI unrelated to the report; 2) general comments about the report; and 3) comments related to specific methodology standards.

Suggestions for PCORI Unrelated to the Report

Some comments specifically address PCORI's mission, policies, and processes. These contributions are extremely valuable to PCORI and have been shared with PCORI leadership. Because these comments lie outside the scope of the PCORI Methodology Report, they are not addressed here.

General Comments about the Report

Comments Expressing Support and Appreciation

Several comments expressed support for the effort to create and disseminate standards for patient-centered outcomes research (PCOR). Most commenters agreed that the standards were aligned with the mission and values of PCORI and that the report and the standards could be used to inform activities of both PCORI and the larger health research community.

Audience and Purpose of the Report

Among commenters, a common theme was that the report's audience was not explicitly stated and seemed to vary greatly among the sections of the draft. The primary audience for the revised report is researchers who conduct PCOR; however, PCORI also recognizes that the research community includes individuals who seek to understand and apply the results of PCOR studies. Report revisions include much new material aimed at making the report accessible to a broader audience. The first measure was to prepare a clear and comprehensive executive summary, which was cited as missing from the draft.

Stories and examples (called Patient Voices, Research Stories, Research in Practice, and CER Wins) have

also been added to place the standards in context and illustrate how good methodology can lead to findings that have a meaningful impact on patient care.

Collaboration and Resources for PCOR

Many commenters identified organizations and individuals as possible collaborators for PCORI in creation of future standards and in PCORI's general work. PCORI is in the process of developing a plan for interaction with the broad community on the creation of new standards. Commenters also provided references and information about resources for conducting PCOR. Where applicable, these sources have been incorporated into the revised report. PCORI thanks the commenters for these suggestions.

Translation Table

A significant number of commenters made reference to the structure and use of the translation table included in the draft report. The importance of such a tool was noted by many, but there were also concerns about the practicality and application of the approach proposed in the draft. Commenters called for:

- revisions to the table that could broadly apply to many study designs and incorporate a broad research context;
- a clear link between the translation table and the standards; and
- instruction as to whether such a table would be a tool used by every applicant.

Guided by these comments, the explanation of the translation table's purpose was revised. The translation table's purpose is to provide a process by which research organizations can evaluate alternative strategies for making trade-offs between the rigor of a research design and the resources (including time) required to carry out specific projects. Thus, the translation table serves a different purpose than the methodology standards. The model for the table has also been simplified to reflect its purpose as an evaluative, rather than prescriptive, tool. This resulted in a broader tool that can be applied to a variety of research designs and settings. The foundation of the table is the framework included in Section II.

Trust Issue and Conflict of Interest

There were a number of issues raised by commenters related to trust and conflict of interest, among them the lack of patient trust for payers, physicians, and industry, and the perception by some

stakeholders that they will be excluded from PCORI's work. Some reviewers felt that the draft report cited examples of limitations and past failings of CER and PCOR that targeted particular actors. PCORI's primary goal in citing examples was not to single out any particular sector, but rather to address the need for broad standards for conflict of interest across the board. Section I of the report has been extensively revised to reflect this goal. Many of the previous examples have been dropped or replaced because they are not germane to the task of developing and disseminating consensus standards for conducting PCOR.

Commenters also asked for more clarification around when researcher and stakeholder affiliations would disqualify them from PCORI-funded research opportunities and when affiliations discount authenticity and the scientific validity of research. This issue was addressed through the revised approach in Section I that removed several references to purported bias in past research. As defined by PCORI's establishing legislation, a conflict of interest is "an association, including a financial or personal association, that have [sic] the potential to bias or have the appearance of biasing an individual's decisions in matters related to the Institute or the conduct of activities." To protect against such conflicts, PCORI will use a strict process for the disclosure of potential conflicts of interest by both researchers and the individuals who serve as reviewers of PCORI funding applications.

Cost

Some submitted comments called upon PCORI and the PCORI Methodology Committee to address in the report several perspectives on cost. Among them were the identification of cost as a significant barrier to PCOR, the high cost associated with some interventions, and the overall burden of cost on the healthcare community. Consistent with its establishing legislation, PCORI does not seek to fund cost-effectiveness research, and the standards do not address cost-effectiveness methodologies.

Dissemination and Implementation of the Report and Standards

A number of commenters made suggestions about potential methods and audiences for dissemination of the report and standards. PCORI and the PCORI Methodology Committee have used these suggestions to guide its comprehensive plan for disseminating and facilitating adoption and implementation of the standards.

The approach to research dissemination that PCORI is developing will address the full range of stakeholders—including consumers and patients; organizations involved in delivering and financing the full range of health-related services, including health-related social services; and policy organizations.

Clarity in Organization, Language, and Definitions

Many commenters made general comments about the organization and terminology included in the report. The revised report was restructured in order to provide a more logical flow, and the Executive Summary now serves as a roadmap for the entire report. Other comments called for definitions to be clarified and, in some cases, even offered revised definitions (examples include PCOR, CER, patients, and stakeholders). An effort has been made to define terms in the text itself, and more terms will be added and defined in a glossary (in progress). The text has been revised to clarify PCORI's mission and scope of work.

Comments Related to the Methodology Standards

Purpose, Development, Application, and Missing Standards

Many commenters called for clarity around the purpose, development, and implementation of the standards. Regarding the intent of the standards, the draft report did not clarify whether the standards developed were best practices, guidelines, or base-level standards. Each term carries with it significant implications about the evidence base available. The revised report emphasizes that the PCORI Methodology Standards should be regarded as the basic requirements for conducting quality PCOR. However, this does not mean that every standard will apply to every study. In the revised report, the standards have been separated into two main groups: General and Cross-Cutting Standards for PCOR and Standards for Specific Study Designs and Methods.

Regarding development of the standards, some commenters were concerned that the creation process lacked transparency, and they said that the process for future standard creation should be clarified. The PCORI Methodology Committee led the development of the initial slate of standards and was closely engaged with many researcher and stakeholder groups. Additionally, the public comment process provided a platform for incorporating advice. Some commenters noted categories of standards that were missing from the initial slate of topics, and others suggested that PCORI should look to other organizations and guideline-making entities for future standards. The report acknowledges the value of

harmonization between the PCORI Methodology Standards and standards disseminated by other entities. Suggested topics have been reviewed by the PCORI Methodology Committee and incorporated into both short- and long-term planning. The suggested topics and standards list generated from the comments will serve as a planning tool for PCORI.

Regarding the implementation of the standards, commenters expressed concern that the large number of standards would be difficult for researchers to implement and would stifle innovation in research methods. The text of the revised report makes clear that not all of the standards apply to any single research project and should not prevent researchers from adding innovative approaches to individual projects. Adherence to these standards will also ensure that the research PCORI funds will be of the highest quality and therefore will have the greatest potential to improve the field of PCOR. PCORI is developing tools and trainings that will help applicants incorporate the standards into their research plans and will help reviewers evaluate the study design and methods of funding applications.

Additionally, PCORI will continue to fund innovative projects, particularly around innovative methods for PCOR. Through the CER Methods program, PCORI seeks to expand and improve the field of research methods. The standards are not viewed as fixed; revising and adding to them will be an ongoing activity for PCORI.

There was very little disagreement about the inclusion of specific standards in the initial slate; however, some commenters provided specific recommendations for individual standards or groups of standards. These submissions, in addition to input from PCORI staff, influenced revisions to the standards. The revised standards were released to the public in late 2012 are the basis of this report. The discussion below concerns the revisions to the text that explains each standard. In addition to these revisions, the following general changes were made to the organization and presentation of all the standards:

- In Section III, the standards were reorganized into the cross-cutting standards (which apply to all studies) and the standards that apply only to studies with particular designs and methods. This was a response to commenters who expressed confusion as to whether all the standards applied to all research and whether the standards were best practices (aspirational) or requirements.
- Each standard is presented only once in the body of the report, and all of the standards are presented together in **Appendix A: PCORI Methodology Standards**. This was done in order to respond to comments that the organization of the draft report was confusing.

Changes to Sections of the Report Providing Rationale for the Methodology Standards

1. Formulating Research Questions

Reviewers noted that standards for formulating research questions are based on current best practices for planning research in general. However, the comments also addressed both research prioritization (i.e., selecting topics to be studied) and specifics of developing research questions once a topic is selected. Research prioritization is now addressed in a separate section of the report that is unrelated to the standards.

The text clarifies that standards for formulating research questions are included to promote transparency and high-quality PCOR research, while acknowledging that these are applicable not only to PCOR. However, an important element of a PCOR program is ensuring that both the research questions and study designs lead to results that are important to patients and relevant to actual patient decisions. This level of specificity and detail makes it easier for individuals and organizations that fund and use research to verify that the research is truly patient-centered.

The changes to the draft standards for this section were minimal. They included: 1) adding to Standard RQ-1 a cross-reference to the new standard on systematic reviews (Standard SR-1); 2) adding data feasibility and quality to the list of elements that should be covered in a formal study protocol in Standard RQ-2; and 3) changing the wording in Standard RQ-5 to broaden who should care about the study outcomes to “people representing the population of interest,” so that this language corresponds to the scope delineated in the patient-centeredness standards.

2. Patient-Centeredness

Many public comments were related to the concept of patient-centeredness in general and how it relates to PCORI activities. These comments informed Sections I and III of the revised report, as appropriate. The standards associated with patient-centeredness as well as the text of this portion of the report were revised to emphasize the significance of these standards to the overall mission of PCORI and high-quality PCOR. The report text acknowledges that the state of the science about best practices and the impact of patient and stakeholder engagement are currently limited, and the text states that the intention of these standards is to promote both patient-centered research and the identification of

optimal approaches to engagement. The text now clarifies that a broad definition of engagement was intended and that individual researchers are not solely responsible for dissemination of findings.

In Standard PC-1, the wording about who should be engaged was revised to read “people representing the population of interest and other relevant stakeholders,” and a description was added that clarifies that this is meant to be inclusive of individuals who have the condition or who are at risk of developing the condition and their surrogates or caregivers, and it may include clinicians, administrators, policy makers, or others involved in healthcare decisions. A related change is an addition to Standard PC-3 to clarify that caregivers may provide “patient-reported outcomes” in cases where patients cannot report for themselves.

Standard PC-4 was rewritten to state that researchers should support, rather than carry out, the dissemination of study results. This change responds to concerns that the draft standard obligated individual researchers to take on dissemination tasks that may not be within their scope of expertise or control. Additionally, the comments made the point that dissemination is often more appropriate and useful when it communicates the findings and implications of a body of research rather than the results of an individual study; therefore, individual researchers can contribute to and support dissemination, although they may not be able to implement a broad dissemination plan based on their individual studies.

3. Data Integrity and Rigorous Analyses

The Data Integrity and Rigorous Analysis standards were named General and Cross-Cutting standards in the draft report. They focus on decisions that are important once a research question has been developed, such as data formats, measurement, and approaches to the assessment of the quality and relevance of individual studies.

The text has been revised to explain that these standards require documentation of key decisions and testing of analysis assumptions as part of the application of any specific study design. The emphasis on data quality, rigor in analyses, and completeness in reporting is not unique to PCOR, but requiring these elements is necessary if PCOR is to be high quality.

A standard originally included in the Data Registry standards was revised to be more general and reclassified as Standard IR-2: Describe data linkage plans, if applicable. This change responded to the observation that data linkages may be important in any studies that use multiple data sources and that concerns should not be limited only to registries. The remaining Data Integrity and Rigorous Analyses standards were not revised from the draft versions.

4. Preventing and Handling Missing Data

Comments about the missing-data standards and text varied from pointing out that the text was very technical to asking for more specifics. The revisions sought to address the range of comments while still advancing the need for rigorous approaches to missing data.

The text has been revised to clarify that these standards apply only to data that researchers intended to collect but did not. They do not apply to data that are “missing by design” (where data may not be available because the study design did not include plans to record certain measurements). The text also now outlines why handling missing data is particularly important for PCOR and therefore appropriate for PCORI Methodology Standards. While the approaches to preventing and handling missing data can be extremely technical, the text has been revised to outline the issues and related standards in terms appropriate for an audience interested in research but not necessarily expert in research design or statistics. The text is also intended to put the standards in context but not prescribe exactly what should be done in any specific study, as this could limit innovation.

Revisions to the standards addressed several public comments. The requirement that researchers should explain their approach to missing data in their analyses and be clear about the implications that their approach may have for the results was added to Standard MD-2. Standard MD-3 was revised so that it is clear that a variety of analytic approaches to missing data may be used, while still reinforcing the idea that single imputation methods are not currently considered best practices and anyone using these methods must provide a scientific rationale for their use. Standard MD-4 was augmented in response to comments that noted that the reasons for patient drop-out are often not recorded or reported in enough detail to allow an assessment of the potential impact that missing data may have on the validity of the study. The standard requires detailed record keeping, accounting for all participants, and continuing data collection unless patient consent is withdrawn.

The wording in Standard MD-5 was revised to allow diverse approaches to the presentation of data for assessing the sensitivity of the results to the methods used to address missing data.

5. Heterogeneity of Treatment Effect

The comments on the standards related to Heterogeneity of Treatment Effect (HTE) generally supported this topic as one of particular importance to PCORI's mission. HTE helps address the core question of which treatments work best for which patients. Commenters correctly pointed out that the standards focus on only one type of approach to HTE—subgroup analyses—and that other approaches that focus on prediction models of benefits to a specific person are available, valid, and becoming more commonly used. The report text has been revised to acknowledge these types of approaches and state that they may be the subject of future standards.

Many public comments suggested topics for future standards or activities, including:

- Specify baseline variables that are particularly important for HTE;
- Provide more specific guidance regarding biomarkers in genetic testing as a basis for HTE analyses;
- Specify the level of evidence needed to claim HTE;
- Develop a checklist for HTE analyses, incorporation of information from individual crossover and n-of-one studies;
- Incorporate HTE analyses into systematic reviews; and
- Extend standards to design, as well as analysis, features of studies that focus on HTE.

PCORI is maintaining lists of potential future topics and activities.

The remainder of the comments raised technical issues about the language of the standards, use of unfamiliar labels or categories, and the omission of some specific analytic. For example, some commenters wanted explicit discussion of quantitative versus qualitative interactions, while another suggested emphasis on simulation and graphical methods.

The revisions to the standards sought to address these technical points and make the standards more broadly inclusive of the analytic approaches currently available. However, some suggestions for stricter rules could inhibit analyses or methodological innovation, and PCORI has tried to avoid creating standards that are overly prescriptive.

Specific revisions to the standards include:

- *Confirmatory, exploratory, and descriptive* were removed as labels of distinction types of subgroup analyses. Standards HT-1 and HT-2 now refer more generally to HTE that is hypothesis driven and hypothesis generating. These initial categories are not yet commonly used in the field, and people found them confusing or felt that they required too much unnecessary explanation. The terms *hypothesis-driven* and *hypothesis-generating* replaced these initial terms.
- The language of Standard HT-4 was revised, and a standard requiring the reporting of statistical power was deleted in response to comments that there was an overemphasis on statistical testing in the standards.
- Language was revised to clarify that HTE is not limited to subgroup effects defined by various baseline covariates but can be done using multivariate scores or classification rules, and this language was added to the standards. The array of statistical approaches to characterizing heterogeneity also was extended to go beyond just testing for interactions and including both estimation and Bayesian approaches in Standards HT-3 and HT-4.

6. Data Registries and 7. Data Networks as Research-Facilitating Infrastructures

The standards for data registries and data networks are similar in that they address both the design and creation of these types of complex data sources for research and also the use of these data in PCOR.

These activities may be undertaken separately by different researchers and organizations or by the same group. Public comments emphasized that these address both infrastructure creation and use.

Suggestions ranged from merging the two into one set of standards to placing them in the same section of the report.

In response, the final report was reorganized so that the standards for data networks immediately follow the standards for data registries. The standards for these two topics now have a parallel construction, thereby further reinforcing the similarity.

Commenters also asked for examples of best practices or further guidance on specific issues such as data quality assurance and privacy protection. These suggestions were judged to be beyond the scope of the current report. Recommendations to endorse a specific system for standardizing terminology were considered premature given the current state of the field.

Several comments raised issues related to the use of electronic medical records for PCOR. This is another key area for future work, which may lead to additional standards in subsequent PCORI methodology reports. Other comments suggested the use of specific data sources or specific skills or

knowledge for researchers. These are unlikely to be included in future standards but may be incorporated into recommendations for research, policy, or PCORI activities in the future.

The Standards DR-1 (for data registries) and DN-1 (for data networks) now list the core requirements for the design of registries and networks as subsections of these overarching standards. Some subsections were revised slightly to increase clarity about the activity covered by the standard and who/which entity was likely to be responsible for that activity. For example, the section of the registry standards on data safety and security (Standard DR-1 B) was revised to clarify that the creators of the registry should provide information to institutional review boards, as these boards are responsible for evaluating data security procedures and assessing risks to patients.

Standard DR-2, which requires researchers to ensure that data registries that are used in their research meet the requirements in Standard DR-1, was not changed, but a similar Standard DN-2 was added to the data network standards. The standards for data registries include a standard about confounding factors (Standard DR-3), which was not changed from the draft report.

8. Causal Inference Methods

Comments on the standards for causal inference methods included several suggestions on how the standards could be revised to promote studies with a strong emphasis on rigorous approaches to causal inference while maintaining a broad scope that would be inclusive of different research methods, study designs, and topics. Several suggestions included proposed changes in the current wording of the standards. Many of the comments were recommendations that additional analytic methods and approaches be included in the standards, as these other methods could contribute to assessing causality in PCOR studies. Commenters also suggested that more detailed standards be included about the two analytic methods specifically mentioned in the standards (propensity scores and instrumental variables).

Creating standards for all analytic methods was not possible in the time frame of developing this initial set of methodology standards. Delineating all possible analytic methods was not possible. Other suggested methods mentioned in comments have been added to the PCORI Methodology Committee's list of potential topics for future standards. Also, although other methods, such as marginal structure models and interrupted time series, are not the subject of specific standards, this does not imply that

they cannot be used in PCORI-sponsored research, and the report text has been revised to clarify this view. Initial, general standards for propensity scores and instrumental variables were included, as these methods are increasingly used in PCOR.

Revisions made to these standards included:

- The term *intervention* was replaced with *exposure* in Standard CI-3 and the report text, as *exposure* is a more general term that can include interventions and encompasses a wider variety of research questions and designs.
- The draft standards referred to exposures and covariates in terms that seemed to limit them to a fixed or specific point in time. Standards CI-1 and CI-4 were revised to allow for time-varying covariates.
- Specific changes were made to clarify that the standards on propensity scores (Standard CI-5) and instrumental variables (Standard CI-6) are minimal standards that are designed to require that the underlying assumptions and the validity of these approaches are both considered by researchers and reported, so that users of the research are able to make their own assessments of the appropriateness of the methods and the assumptions and evaluate the utility of the results.

9. Adaptive and Bayesian Trial Designs

Many comments were submitted about the section of the draft report on standards for adaptive and Bayesian trials. However, few comments were related to the standards directly. Most comments raised issues related to the value of randomized controlled trials (RCTs) versus observational studies. These comments also described issues surrounding placebo-controlled trials, elicitation of pretrial treatment or outcome preferences, quasi-experimental designs, how to choose outcome measures, preference trials, cultural competency and trial ethics, how interventions adapt to local contexts, the relevance of RCTs to implementation science, the role of modeling as an alternative to trials, and the need to expand the range of designs covered by the standards (e.g., include cluster randomized trials).

These fundamental issues about the appropriateness of study designs for PCOR and how designs can be improved and used to promote high-quality research are overarching questions that are driving activities of PCORI as well as other stakeholders. The translation framework that underlies the translation table also addresses these issues.

The only modification made to the standards was in Standard AT-4, which outlines the infrastructure needed to support these types of trials. The revision clarifies that part of this infrastructure includes

having an analytic group that is independent of trial investigators, so that investigators are not apprised of the randomization probabilities, which encode information on treatment effectiveness or safety.

10. Studies of Diagnostic Tests

Most of the comments received about diagnostic tests suggested expansion of the scope of tests and approaches that should be covered by the standards or requested clarification on the intention of the standard (now Standard DT-5) that encourages the use of prospective randomized studies.

Comments suggested additional types of tests, diagnostic challenges, and disease-specific questions that could be addressed by the standards. For this initial group of standards, the scope of the standards excludes studies of predictive accuracy, genetic testing, and modeling. The text of the report was revised to clarify the scope and reinforce that these standards will be revised and augmented in the future.

One standard relating to diagnostic tests, Standard DT-5, was modified to clarify the intent of emphasizing a particular type of research in studies of diagnostic tests. Studies of test outcomes should ideally use a prospective randomized study design, but this does not preclude other designs if they are warranted in a given situation. In assessing the impact of diagnostic tests on patient outcomes, if the most relevant patient-centered outcome cannot be feasibly tested, then a previously validated surrogate should be used. If a nonrandomized design is proposed, then the rationale for using an observational study should be provided, and efforts to minimize confounding should be documented. This standard was created to address one of the most challenging aspects in diagnostic test evaluation. Although prospective randomized designs are most effective in the comparison of the impact of tests on patient outcomes, often such designs are not practically feasible. Realistic alternatives need to be considered, and the standard now points to some of them.

11. Systematic Reviews

As several commenters noted, the Institute of Medicine (IOM) has published standards on systematic reviews, and standards also have been disseminated by the Agency for Healthcare Research and Quality (AHRQ). Commenters also raised issues such as the applicability of standards on systematic reviews to reviews on delivery system interventions (as opposed to reviews on specific medical/surgical interventions).

To address these concerns, Standard SR-1 was added. This standard endorses the IOM standards for conducting systematic reviews of clinical effectiveness but adds that there can be some flexibility in applying them. Text was added to the report to justify the standard and explained that flexibility is justified because there is limited empirical evidence to support some aspects of the IOM standards.

APPENDIX C: RECOMMENDED ACTIONS AND RESEARCH RECOMMENDATIONS

In conjunction with the creation of the draft methodology report and standards, PCORI's Methodology Committee developed a set of recommended actions and research to help PCORI fund high-impact, patient-centered research and accomplish the organization's mission.

Recommended Actions

Listed below are the recommended actions already initiated or under development.

Topic	Initiated	Under Development
Patient-Centeredness	<ul style="list-style-type: none"> • Improve the patient-reported outcomes (PRO) evidence base by supporting research on methods for assessing measurement properties (based on qualitative and quantitative evaluations), score interpretability, meaningfulness of score changes, and strategies for minimizing and interpreting missing PRO data in patient-centered outcomes research (PCOR). • Include patients in the peer review process. • Develop PCORI infrastructure and services to support and facilitate investigators' engagement activities. These services include support for identification and recruitment of patient representatives and for the selected representatives' (and other stakeholders') involvement in PCORI-funded research. • Develop a sample patient engagement plan to demonstrate the key elements required for patient engagement in the research process. The sample plan will illustrate engagement of both patient representatives and study participants to facilitate investigator adherence to PCORI standards. 	<ul style="list-style-type: none"> • Evaluate the effectiveness of patient dissemination activities.
	<ul style="list-style-type: none"> • Establish peer review processes that avoid interference from participants 	

Research Prioritization	<p>and stakeholders with potential conflicts of interest.</p> <ul style="list-style-type: none"> • Incorporate patient and stakeholder perspectives in ongoing evaluation of peer review. • Protect independence in peer review of research funding proposals. • Adopt methods of peer review to safeguard independence between reviewers and those being reviewed. • Ensure adequate representation of minorities and disadvantaged segments of the population in peer review of research funding proposals. • In non-targeted funding opportunity announcements, require applicants to demonstrate how their proposed research fills a research gap. • Implement a process by which the public is encouraged to nominate topics for new research. • Base all PCORI targeted-funding announcements on evidence gap analysis. 	
Dissemination	<ul style="list-style-type: none"> • In collaboration with external partners, develop infrastructure to realize the efficiency inherent in centralized dissemination of research findings and address limitations in researchers' capacity and ability to disseminate findings to the full range of relevant stakeholders. 	
Training and Implementation of the Standards	<ul style="list-style-type: none"> • Develop training and tools to guide applicants and reviewers in the preparation and review of proposals with respect to key elements of the standards. Subjects to be included in training: <ul style="list-style-type: none"> ○ Patient engagement methods for investigators; ○ Value-of-information (VOI) methodologies; ○ Systematic reviews; ○ Modeling; ○ Addressing missing data; ○ Controlling for confounding; ○ Adaptive trials; and ○ Research dissemination methods for investigators. 	

Data Integrity and Rigorous Analysis		<ul style="list-style-type: none"> • Encourage observational components to randomized designs or vice versa and explore other ways to make PCORI-funded research methodologically informative. • Encourage and actively seek studies that evaluate clinical outcomes after diagnostic testing.

Recommended Research

Topic	Recommended Research
Patient-Centeredness	<ul style="list-style-type: none"> • Support research to develop a standardized nomenclature for patient engagement methods. • Encourage studies, ideally with experimental designs, that assess different methods for engaging patients with diverse views and preferences and funneling their input into the peer review process in a consultative manner.
Research Prioritization	<ul style="list-style-type: none"> • Encourage intra- and extramural research in the development and practical application of VOI methods for PCOR, including studies that examine the contribution of VOI methods to research prioritization when used in conjunction with other approaches to research prioritization. • Support empirical research to assess and improve research prioritization methods for use by PCORI. • Support extra- and intramural research to establish a best-practice approach to consultative and collaborative patient engagement in topic generation that is suitable for the heterogeneity of the US patient population.

	<ul style="list-style-type: none"> • Study the employment of research gap analysis to continue to develop the empirical evidence on its use.
Data Integrity and Rigorous Analysis	<ul style="list-style-type: none"> • Fund research on ways to identify and include reasonable treatment alternative comparators. • Develop and disseminate templates for describing who is in each analysis and the potential sources of selection bias. • Develop and disseminate methods for adequate analysis of data in cases where the treatment or exposure varies over time. • Fund research on the best way to harmonize data elements across sources.
Heterogeneity of Treatment Effects (HTE)	<ul style="list-style-type: none"> • Develop methods guidance on the use of Bayesian methods in HTE analyses and an appropriate outcome scale for HTE analysis (e.g., risk difference, risk ratio, log of odds ratio). • Support the development of both analytic approaches and guidance for predictive approaches to HTE. • Develop methods guidance for HTE analyses in comparative effectiveness trials; the literature on HTE almost exclusively discusses use in placebo-controlled trials.
Data Registries	<ul style="list-style-type: none"> • Develop analytic techniques for addressing measured and unmeasured confounding. • Develop analytic techniques for handling missing data that can be used in registry studies. • Develop improved strategies for linking data while maintaining privacy protections. • Develop innovative ways to reduce loss to follow-up as registries encompass longer time periods.

APPENDIX D: THE PCORI METHODOLOGY COMMITTEE'S APPROACH TO THE AUTHORIZING LEGISLATION

PCORI's authorizing legislation¹ directed the PCORI Methodology Committee to pursue three major activities: creation of a "translation table," development of methodological standards for research, and generation of recommendations to promote adherence to the methodological standards. The committee began its work by defining terms and considering the various approaches that might be taken in responding to these specific requirements in the legislation.

Translation table. The legislation specifies that the PCORI Methodology Committee must develop or devise "a translation table that is designed to provide guidance and act as a reference for the Board to determine research methods that are most likely to address each specific research question." Section II of this report provides a general framework for the translation table. The methodology committee is committed to developing a translation framework that provides a structure for creating translation tables that can be used by various stakeholder groups (including the PCORI Board of Governors). The framework provides a mechanism for defining the trade-offs and decisions involved in selecting a study design.

Methodological standards for research. The language of the legislation around the establishment of methodological standards has several salient features that outline the Congressional expectations for the committee and its work. First, the legislation defines expectations with regard to the *process* by which the committee should proceed, stipulating that the establishment of methodological standards should, as appropriate, build on existing work, and that the process should be ongoing, scientifically based, and inclusive. Second, the legislation specifies both specific and general *content* areas that the proposed standards should address, including internal validity, generalizability, feasibility, and timeliness of research. Third, the standards should provide specific criteria for health outcome measures, risk adjustment, and other relevant aspects of research and assessment with respect to the design of research. Fourth, standards must include methods that account for and evaluate different patient subpopulations in various types of research. Fifth, development of the standards must include input from stakeholders and the public. (This requirement was met by soliciting public comment on PCORI's initial draft set of methodology standards. The standards were revised and updated in response to the comments submitted by a broad range of stakeholders.) Finally the *scope* of the committee's work is to include "each of the major categories of comparative clinical effectiveness research methods," which are listed as:

- "Systematic reviews and assessments of existing and future research and evidence including original research conducted subsequent to the date of the enactment of this section."

¹ Available at pcori.org/assets/PCORI-Authorizing-Legislation-032310.pdf

- Primary research, such as randomized clinical trials, molecularly informed trials, and observational studies.
- Any other methodologies recommended by the methodology committee established under paragraph (6) that are adopted by the Board under paragraph (9)."

Recommended actions to comply with methodological standards. The enabling legislation instructs the committee not only to propose methodological standards but also to recommend actions necessary to comply with them: "The methodology committee shall submit reports to the Board.... Reports shall contain recommendations for the Institute to adopt methodological standards ... as well as other actions deemed necessary to comply with such methodological standards."

Development of the Translation Framework

As the basis for developing a translation framework that can guide the creation of translation tables, PCORI staff searched MEDLINE, Scopus, and the [AHRQ Scientific Resource Center Methods Database](#)² for articles that use the term *translation table*. A Request for Information was also used to gather stakeholder input regarding the translation table. The draft framework was released for public comments between July 23 and September 14, 2012. The translation framework was revised based on these comments and was finalized through a consensus process. It is described in Section II of this report.

Defining Methodological Standards

With a legislative direction for the process, content, and scope of the standards, the PCORI Methodology Committee began work on developing an initial slate of standards. The Committee initially established priorities for the core set of essential standards. The Committee formed small working groups to explore patient-centeredness, research prioritization, and research methods. A fourth group coordinated communication and prepared draft report components. The goal was to create methodological standards in important research domains that are representative of research issues in comparative effectiveness research. This process allowed the committee to complete the inventory of recommended standards and to establish a process for periodically reviewing and updating them.

² Available at citeulike.org/user/SRCMethodsLibrary

Building on the work of the Institute of Medicine (IOM 2011), the Committee defined a standard as follows:

- A process, action, or procedure for performing patient-centered outcomes research (PCOR) that is deemed essential to producing scientifically valid, transparent, and reproducible results.
- A standard may be supported by scientific evidence, reasonable expectation that the standard helps achieve the anticipated level of quality in PCOR, or broad acceptance of the practice in PCOR.
- A standard is actionable, feasible, and implementable.

Select and Assess Proposed Standards

The working groups developed provisional lists of major research method topics and chose 129 topics for focused review. Contractors were secured to assist the PCORI Methodology Committee in developing materials for each topic. Full reports are available on [the PCORI website](#).³ In addition to the full reports, contractors summarized key information regarding each proposed standard in a template format to assist committee members in making comparisons among the proposed standards, based broadly on criteria derived from [AGREE](#)⁴, an international project developing guidelines for the appraisal of research and evaluation. Four criteria were deemed especially important: contribution to patient-centeredness, contribution to scientific rigor, contribution to transparency, and empirical evidence/theoretical basis. (See **Criteria for Adopting Proposed Standards.**)

The working groups held workshops at which contractors presented their findings and recommendations for discussion with committee members, invited experts, and patient representatives. This process reduced the original list of 129 topics to 88 preliminary standards by eliminating those that were out of scope or redundant.

³ Available at pcori.org/research-we-support/methodology/about-the-methodology-committee

⁴ Available at agreertrust.org/wp-content/uploads/2013/06/AGREE_II_Users_Manual_and_23-item_Instrument_ENGLISH.pdf

The committee then conducted an in-depth review of the 88 preliminary standards. Committee members were asked to consider each standard as a minimum requirement for PCOR. To become a formal standard, the topic had to be approved by a minimum of two-thirds of members (12 of 17) of the PCORI Methodology Committee.

Criteria for Adopting Proposed Standards

Criterion	Definition
Contribution to patient-centeredness	The degree to which the proposed standard contributes to respect for and responsiveness to individual patient preferences, needs, and values; whether the proposed standard would help ensure that patient values and circumstances guide clinical decisions.
Contribution to scientific rigor	The degree to which the proposed standard contributes to objectivity, minimizes bias, improves reproducibility, and leads to more complete reporting.
Contribution to transparency	The degree to which the proposed standard contributes to explicit methods, consistent application, and the opportunity for public review, so that users can link judgments, decisions, or actions to the data on which they are based. The degree to which it allows users to assess the strengths and weaknesses of the study to which the standard is applied.
Empirical evidence and theoretical basis	Description of the information upon which a proposed standard is based, emphasizing empirical evidence about the proposed standard and theoretical support.
Degree of controversy about use of standard	Description of controversy or alternative views of the proposed standard, particularly with respect to criteria above (e.g. patient-centeredness, scientific rigor, transparency) in the context of comparative effectiveness research.
Other considerations	Description of other considerations that might influence adoption of the proposed standard, such as practicality, feasibility, barriers to implementation, and cost.

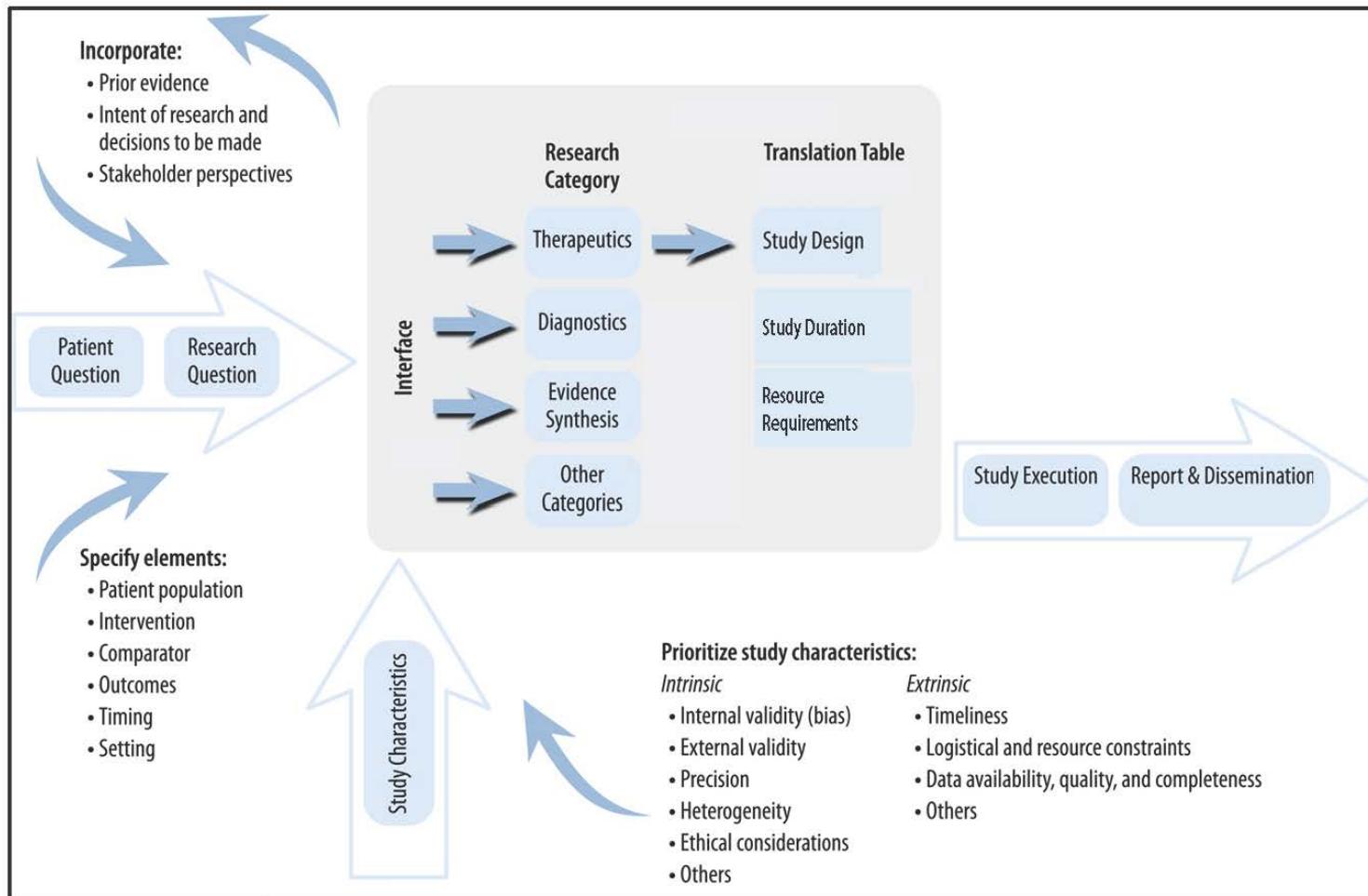
The draft methodology standards were released for public comment between July 23 and September 14, 2012. PCORI received, collated, and summarized 1487 unique comments. Small working groups within the methodology committee carefully reviewed each standard in the

context of comments from the PCORI Board of Governors and from the public, and the committee revised the standards using a consensus process. The final PCORI Methodology Standards were published December 14, 2012, after the Board accepted them in a public session on November 19, 2012.

Recommended Actions to Comply with Methodological Standards

The PCORI Methodology Committee, Board, and staff are developing a coordinated approach to broad dissemination and implementation of the standards. This includes engaging stakeholders who might use the standards, creating reporting and surveillance opportunities, and creating tools and training to increase compliance with the standards and ability to evaluate incorporation of the standards into the peer review of applications for PCORI funding. Other actions recommended by the PCORI Methodology Committee to promote adherence to the standards are outlined in **Appendix C: Recommended Actions and Research Recommendations**.

APPENDIX E: TRANSLATION FRAMEWORK



APPENDIX F: REFERENCES

- Blackwood, B., F. Alderdice, K. Burns, C. Cardwell, G. Lavery, and P. O'Halloran. 2011. "Use of Weaning Protocols for Reducing Duration of Mechanical Ventilation in Critically Ill Adult Patients: Cochrane Systematic Review and Meta-analysis." *BMJ* 342:c7237.
- Bossuyt, P.M., and K. McCaffery. 2009. "Additional Patient Outcomes and Pathways in Evaluations of Testing." *Med Decis Making* 29(5):E30–8.
- Bossuyt, P.M., J.B. Reitsma, D.E. Bruns, C.A. Gatsonis, P.P. Glasziou, L.M. Irwig, et al. 2003a. "Towards Complete and Accurate Reporting of Studies of Diagnostic Accuracy: The STARD Initiative." *BMJ* 326(7379):41–4.
- Bossuyt, P.M., J.B. Reitsma, D.E. Bruns, C.A. Gatsonis, P.P. Glasziou, L.M. Irwig, et al. 2003b. "The STARD Statement for Reporting Studies of Diagnostic Accuracy: Explanation and Elaboration." *Clin Chem* 49(1):7–18.
- Bravata, D.M., K.M. McDonald, K.G. Shojania, V. Sundaram, and D K. Owens. 2005. "Challenges in Systematic Reviews: Synthesis of Topics Related to the Delivery, Organization, and Financing of Health Care." *Ann Intern Med* 14(12 Pt 2):1056–65.
- Brookes, S.T., E. Whitley, T.J. Peters, P.A. Mulheran, M. Egger, and G. Davey Smith. 2001. "Subgroup Analyses in Randomised Controlled Trials: Quantifying the Risks of False-positives and False-negatives." *Health Technol Assess* 5(33):1–56.
- Buede, D. 2005. "Influence Diagrams: A Practitioner's Perspective." *Decision Analysis* 2(4):235–7.
- CAST-II (Cardiac Arrhythmia Suppression Trial-II) Investigators. 1992. "Effect of Antiarrhythmic Agent Moricizine on Survival after Myocardial Infarction: The Cardiac Arrhythmia Suppression Trial-II." *N Engl J Med* 327:227–33.
- Carlson, J.J., et al. 2013. "Value-of-Information Analysis within a Stakeholder-driven Research Prioritization Process in a US Setting: An Application in Cancer Genomics." *Med Decis Making* 33(4):463–71.
- Claxton, K.P., and M.J. Sculpher. 2006. "Using Value of Information Analysis to Prioritise Health Research: Some Lessons from Recent UK Experience." *Pharmacoeconomics* 24(11):1055–68.
- Crandall, W.V., P.A. Margolis, M.D. Kappelman, E.C. King, J.M. Pratt, B.M. Boyle, L.F. Duffy, J.E. Grunow, S.C. Kim, I. Leibowitz, B.T. Schoen, R.B. Colletti, ImproveCareNow Collaborative. 2012. "Improved outcomes in a quality improvement collaborative for pediatric inflammatory bowel disease." *Pediatrics* 129:e1030-41.
- Detre, K., et al. 1981. "Effect of Bypass Surgery on Survival in Patients in Low- and High-risk Subgroups Delineated by the Use of Simple Clinical Variables." *Circulation* 163:1329–38.
- Deverka, P.A., S.D. Schully, N. Ishibe, J.J. Carlson, A. Freedman, K.A. Goddard, et al. 2012. "Stakeholder Assessment of the Evidence for Cancer Genomic Tests: Insights from Three Case Studies." *Genet Med* 14: 656–62.
- Echt, D.S., P.R. Liebson, L.B. Mitchell, et al. 1991. "Mortality and Morbidity in Patients Receiving Encainide, Flecainide, or Placebo—The Cardiac Arrhythmia Suppression Trial." *N Engl J Med* 324:781–8.
- Ferrante di Ruffano, L., C.J. Hyde, K.J. McCaffery, P.M. Bossuyt, and J.J. Deeks. 2012. "Assessing the Value of Diagnostic Tests: A Framework for Designing and Evaluating Trials." *BMJ* 344:e686.
- Fiore, L.D., M. Brophy, R.E. Ferguson, L. D'Avolio, J.A. Hermos, R.A. Lew, et al. 2011. "A Point-of-Care Clinical Trial Comparing Insulin Administered Using a Sliding Scale Versus a Weight-based Regimen." *Clin Trials* 8(2):183–95.
- Food and Drug Administration. 2010a. "Guidance for Industry: Adaptive Design Clinical Trials for Drugs and Biologics." fda.gov/downloads/DrugsGuidanceComplianceRegulatoryInformation/Guidances/UCM201790.pdf.

- Food and Drug Administration. 2010b. "Guidance for Industry and FDA Staff: Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials." fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf.
- Food and Drug Administration. 2007. "Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests." fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071148.htm.
- Gabriel, S.E., and S.L.T. Normand. 2012. "Getting the Methods Right—The Foundation of Patient-Centered Outcomes Research." *NEJM* 367(9):787-9.
- Gagnon, M. P., M. Desmartis, D. Lepage-Savary, J. Gagnon, M. St-Pierre, M. Rhainds, et al. 2011. "Introducing Patients' and the Public's Perspectives to Health Technology Assessment: A Systematic Review of International Experiences." *Int J Technol Assess Health Care* 27(1):31–42.
- Gliklich, R.E., M.B. Leavy, P. Velentgas, N.A. Dreyer, S.R. Tunis, P.Mohr, D.A. Messner, R.M. Moloney, S.U. Karkare, R.W. Dubois, and J.S. Graff. 2012. "Incorporating Stakeholder Perspectives in Developing a Translation Table Framework for Comparative Effectiveness Research." *J Comp Eff Res* 1 (3): 281–92.
- Goldfine, A.B., S. Kaul, and W.R. Hiatt. 2011. "Fibrates in the Treatment of Dyslipidemias—Time for a Reassessment." *N Engl J Med* 365(6):481–4.
- Guyatt, G., D. Rennie, M. Meade, and D. Cook. 2008. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. 2nd Edition. New York: The McGraw-Hill Companies, Inc.
- Helfand, M., S. Tunis, E.P. Whitlock, S.G. Pauker, A. Basu, J. Chilingerian, et al. 2011. "A CTSA Agenda to Advance Methods for Comparative Effectiveness Research." *Clin Transl Sci* 4(3):188–98.
- Hernán, M., and R. McKaig. 2011. "The HIV-CAUSAL Collaboration; When to Initiate Combined Antiretroviral Therapy to Reduce Mortality and AIDS-Defining Illness in HIV-Infected Persons in Developed Countries. An Observational Study." *Ann Int Med* 154(8):509–15. hsph.harvard.edu/miguel-hernan/hiv-causal-collaboration.
- Hess, E.P., M.A. Knoedler, N.D. Shah, J.A. Kline, M. Breslin, M.E. Branda, L.J. Pencille, B.R. Asplin, D.M. Nestler, A.T. Sadosty, I.G. Siell, H.H. Ting, and V.M. Montori. 2012. "The Chest Pain Choice Decision Aid: A Randomized Trial." *Circ Cardiovasc Qual Outcomes* 5:251-9.
- Institute of Medicine (IOM). 2008. "Knowing What Works in Health Care: A Roadmap for the Nation." Consensus Report. Cited January 10, 2012. iom.edu/Reports/2008/Knowing-What-Works-in-Health-Care-A-Roadmap-for-the-Nation.aspx.
- . 2009. *Initial National Priorities for Comparative Effectiveness Research*. Washington, DC: The National Academies Press. nap.edu/openbook.php?record_id=12648&page=29.
- . 2011. *Finding What Works in Health Care: Standards for Systematic Reviews*. Washington, DC: The National Academies Press.
- ISIS-1 (First International Study of Infarct Survival) Collaborative Group. 1986. "Randomized Trial of Intravenous Atenolol among 16027 Cases of Suspected Acute Myocardial Infarction: ISIS-1." *Lancet* 2(8498):57–66.
- Kim, L.G., A.P. Scott, et al. 2007. "A Sustained Mortality Benefit from Screening for Abdominal Aortic Aneurysm." *Ann Int Med* 146:699–706.
- Kotchen,T.A., and R. Spelley 2012. "Peer Review: A Research Priority." White Paper. pcori.org/assets/Peer-Review-A-Research-Priority.pdf.
- Krumholz, H. 2012. "Missing Data: The Elephant That's Not in the Room." cardioexchange.org/voices/missing-data-the-elephant-thats-not-in-the-room/
- Lagakos, S.W. 2006. "The Challenge of Subgroup Analyses—Reporting without Distorting." *N Engl J Med* 354(16):1667–9.

- Lipitz-Snyderman, A., D. Steinwachs, D M. Needham, E. Colantuoni , L.L. Morlock, and P.J. Pronovost. 2011. "Impact of a Statewide Intensive-Care Unit Quality-Improvement Initiative on Hospital Mortality and Length of Stay: Retrospective Comparative Analysis." *BMJ* 342(Jan 28): d219. doi: 10.1136/bmj.d219.
ncbi.nlm.nih.gov/pmc/articles/PMC3031651/pdf/bmj.d219.pdf
- Lohr, K.N. 2007. "Emerging Methods in Comparative Effectiveness and Safety: Symposium Overview and Summary." *Med Care* 45(10 Supl 2):S5–8.
- Lord, S.J., L. Irwig, and P.M.M. Bossuyt. 2009. "Using the Principles of Randomized Controlled Trial Design to Guide Test Evaluation." Medical Tests—White Paper Series. Rockville, MD: Agency for Healthcare Research and Quality (US).
- Matchar, D. 2012. "Introduction to the Methods Guide for Medical Test Reviews." *J Gen Intern Med* 27 Suppl 1:S4–10.
- Meltzer, D.O. 2001. "Addressing Uncertainty in Medical Cost-Effectiveness Analysis Implications of Expected Utility Maximization for Methods to Perform Sensitivity Analysis and the Use of Cost-Effectiveness Analysis to Set Priorities for Medical Research." *J Health Econ* 20(1):109–29.
- , T. Hoomans, J.W. Chung, and A. Basu. 2011. "Minimal Modeling Approaches to Value of Information Analysis for Health Research." *Med Decis Making* 31(6):E1–E22.
- Meyers, E., A.J. McBroom, L. Shen, R.E. Posey, R. Gray, and G. Sanders. 2012. "Value-of-Information for Patient-Centered Outcomes Research Prioritization." White Paper. pcori.org/assets/Value-of-Information-Analysis-for-Patient-Centered-Outcomes-Research-Prioritization.pdf.
- Moher, D., S. Hopewell, K.F. Schulz, V. Montori, P.C. Gotzsche, P.J. Devereaux, et al. 2010. "CONSORT 2010 Explanation and Elaboration: Updated Guidelines for Reporting Parallel Group Randomised Trials." *BMJ* 340:c869.
- Molnar, F.J., et al. 2009. "Have Last-Observation-Carried-Forward Analyses Caused Us to Favour More Toxic Dementia Therapies over Less Toxic Alternatives? A Systematic Review." *Open Med* 3:31–50.
- Montori, V.M., S.P. Kim, G.H. Guyatt, N.D. Shah. 2012. "Which Design for Which Question? An Exploration toward a Translation Table for Comparative Effectiveness Research." *J Comp Eff Res* 1(3):271–9.
- Mullins, C.D., B. Barnet, S. dosReis, K.S. Kauffman, E. Onukwugha. 2012. "Integrating Patients' Voices in Study Design Elements with a Focus on Hard-to-Reach Populations." White Paper.
pcori.org/assets/pdfs/Integrating%20Patients%20Voices.pdf.
- Muss, H B., D.A. Berry, C.T. Cirrincione, M. Theodoulou, A.M. Mauer, A.B. Kornblith, et al. 2009. "Adjuvant Chemotherapy in Older Women with Early-Stage Breast Cancer." *N Engl J Med* 360(20):2055–65.
- Nass, P., S. Levine, and C. Yancy. 2012. "Methods for Involving Patients in Topic Generation for Patient-Centered Comparative Effectiveness Research—An International Perspective." Research Priorities White Paper. Patient-Centered Outcomes Research Institute. pcori.org/assets/Methods-for-Involving-Patients-in-Topic-Generation-for-Patient-Centered-Comparative-Effectiveness-Research-%E2%80%93-An-International-Perspective.pdf
- National Research Council. 2010. The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.
nap.edu/openbook.php?record_id=12955.d
- Oliver, S. R., R. W. Rees, L. Clarke-Jones, R. Milne, A. R. Oakley, J. Gabbay, et al. 2008. "A Multidimensional Conceptual Framework for Analysing Public Involvement in Health Services Research." *Health Expect* 11(1):72–84.

- Organisation for Economic Co-operation and Development (OECD). 2013. "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data." oecd.org/document/18/0,3746,en_2649_34223_1815186_1_1_1,00.html.
- Patient-Centered Outcomes Research Institute (PCORI) Methodology Committee. 2012. "Methodological Standards and Patient-Centeredness in Comparative Effectiveness Research." *JAMA* 307(15):1636–40.
- Patient Partner. 2012. "The Active Involvement of Patients and Patient Organisations in (the) Clinical Trial (Development Process)." patientpartner-europe.eu/en/resources/active-involvement-in-the-process.
- Pinto, D.S., P.D. Frederick, A.K. Chakrabarti, A.J. Kirtane, E. Ullman, A. Dejam, D.P. Miller, T.D. Henry, C.M. Gibson, and National Registry of Myocardial Infarction Investigators. 2011. "Benefit of Transferring ST-Segment-Elevation Myocardial Infarction Patients for Percutaneous Coronary Intervention Compared with Administration of Onsite Fibrinolytic Declines as Delays Increase." *Circulation* 124(23):2512–21.
- Pronovost, P., D. Needham, S. Berenholtz, et al. 2006. "An Intervention to Decrease Catheter-related Bloodstream Infections in the ICU." *N Engl J Med* 355:2725–32.
- Raiffa, H., and R. Schlaifer. 1961. *Applied Statistical Decision Theory*. Harvard Business School, Colonial Press.
- Ransohoff, D. F. 2007. "How to Improve Reliability and Efficiency of Research about Molecular Markers: Roles of Phases, Guidelines, and Study Design." *J Clin Epidemiol* 60(12):1205–19.
- Rein, David. 2012. "Value of Information and Research Prioritization." White Paper. pcori.org/assets/Value-of-Information-and-Research-Prioritization.pdf
- Reitsma, J., A. Rutjes, P. Whiting, V. Vlassov, M. Leeflang , and J. Deeks. 2009. "Assessing Methodological Quality." In *Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 100: The Cochrane Collaboration*, edited by J.J.B.P. Deeks and C. Gatsonis. Chapter 9. London.
- Rosenbaum, P., and D. Rubin. 1984. "Reducing Bias in Observational Studies Using Subclassification on the Propensity Score." *J Am Statist Assoc* 79(387):516–24.
- Rothwell, P.M., G.R. Fowkes, J.F. Belch, H. Ogawa, C.P. Warlow, and T.W. Meade. 2011. "Effect of Daily Aspirin on Long-Term Risk of Death Due to Cancer: Analysis of Individual Patient Data from Randomised Trials." *Lancet* 377(9759):1–41.
- Rothwell, P.M., M. Wilson, J.F. Price, J.F. Belch, T.W. Meade, and Z. Mehta. 2012. "Effect of Daily Aspirin on Risk of Cancer Metastasis: A Study of Incident Cancers During Randomised Controlled Trials." *Lancet* 379(9826): 1591–601
- Santaguida, P., C. Riley, and D. Matchar. 2012. "Chapter 5: Assessing Risk of Bias as a Domain of Quality in Medical Test Studies." *J Gen Intern Med* 27(1 Supplement):33–8.
- Schneeweiss, S., J.D. Seeger, and S.R. Smith. 2012. "Methods for Developing and Analyzing Clinically Rich Data for Patient-Centered Outcomes Research: An Overview." *Pharmacoepidem Drug Safe* 21(S2):1–5.
- Schulz, K.F., D.G. Altman, and D. Moher. 2010. "CONSORT 2010 Statement: Updated Guidelines for Reporting Parallel Group Randomized Trials." *Ann Intern Med* 152(11):726–32.
- Smith, A.J., et al. 2012. "Failure Rates of Stemmed Metal-on-Metal Hip Replacements: Analysis of Data from the National Joint Registry of England and Wales." *Lancet* 379:1199–204.
- Sox, H.C., and S.N. Goodman. 2012. "The Methods of Comparative Effectiveness Research." *Annual Rev Public Health* 33:425–45.
- Staniszewska, S., J. Brett, C. Mockford, and R. Barber. 2011. "The GRIPP Checklist: Strengthening the Quality of Patient and Public Involvement Reporting in Research." *Int J Technol Assess Health Care* 27(4):391–9.
- Tunis, S.R., D.A. Messner, P. Mohr, R.E. Gliklich, and R.W. Dubois. 2012. "A Translation Table for Patient-Centered Comparative Effectiveness Research: Guidance to Improve the Value of Research for Clinical and Health Policy Decision-Making." *J Comp Eff Res* 1(3):259–62.

- Trikalinos, T., C. Balion, and C. Coleman. 2012. "Chapter 8: Meta-Analysis of Test Performance When There Is a 'Gold Standard.'" *J Gen Intern Med* 27 (1 Supplement):56–66.
- Trikalinos, T., and C. Balion. 2012. "Chapter 9: Options for Summarizing Medical Test Performance in the Absence of a 'Gold Standard.'" *J Gen Intern Med* 27:1 Supplement):67–75.
- Trikalinos, T., S. Kulasingam, and W. Lawrence. 2012. "Chapter 10: Deciding Whether to Complement a Systematic Review of Medical Tests with Decision Modeling." *J Gen Intern Med* 27:(1 Supplement):75–82.
- Whiting, P.F., A.W. Rutjes, M.E. Westwood, S. Mallett, J.J. Deeks, J.B. Reitsma, et al. 2011. "QUADAS-2: A Revised Tool for the Quality Assessment of Diagnostic Accuracy Studies." *Ann Intern Med* 155(8):529–36.
- Whiting, P.F., M.E. Weswood, A.W. Rutjes J.B. Reitsma, P.N. Bossuyt, and J. Kleijnen. 2006. "Evaluation of QUADAS, a Tool for the Quality Assessment of Diagnostic Accuracy Studies." *BMC Med Res Methodol* 6:9.

APPENDIX G: CONTRIBUTORS

The editors and the PCORI Methodology Committee would like to offer special thanks to the individuals who took part in revision and review of this document, including Susan Hildebrandt, Stanley Ip, Maggi LeDuc, Sally Morton, Sandi Myers, and Sue Sheridan.

We would also like to recognize the individuals who supported the development of the stories and examples, including Josh Carlson, Michael Demers, Jacqueline Fridge, Eric Hess, Annie LeBlanc, Michel LeBlanc, Courtney Schreiber, Lucinda Shore, Leigh Simmons, Beryl, Juli, Steve, and Suzanne. Kay Dickersin provided assistance in identifying sources for stories. We also thank the patients, researchers, and members of the PCORI Methodology Committee and Board of Governors who reviewed earlier drafts of the stories.

Additionally we thank the following organizations for their contribution to the stories and examples, including [Informed Decision Making Foundation](#)¹ for *Patient Voices: Juli*, on page 31; and [Healthtalkonline.org](#)² for *Patient Voices: Sarah*, on page 40. Healthtalkonline.org is a website run by the DIPEx Charity, and this example is based on research led by the Health Experiences Research Group, Department of Primary Care Health Sciences, University of Oxford.

In addition, the editors and the PCORI Methodology Committee wish to thank individuals who were instrumental in creating the draft methodology report in 2012. Editors of the draft report included Mark Helfand, Alfred Berg, David Flum, Sherine Gabriel, and Sharon-Lise Normand. Former committee member John Ioannidis provided assistance in drafting portions of the draft report. The committee is especially indebted to former PCORI consultant, Gail Shearer, who provided support for the committee and the report. We also wish to thank members of the draft report editorial team, including Andrew Holtz, Heidi Nelson, Edwin Reid, Lori Frank, Camber Hansen-Karr, Lynn Kitagawa, Lauren Saxton, and Leah Williams and members of the research team from Deloitte Consulting LLC, including Constanze Coon, Anna Cottone, Milianne Ly, and Rayneisha Watson.

¹ Available at <http://www.informedmedicaldecisions.org>

² Available at <http://www.healthtalkonline.org>

The committee offers particular thanks to the individuals and organizations who were commissioned to create the reports that influenced development of the Methodology Standards. These reports are available on [PCORI's website](#).³

³ Available at pcori.org/research-we-support/methodology/about-the-methodology-committee