PCORI METHODOLOGY REPORT

January 2019

The Methodology Standards were updated in February 2019 and are available at www.pcori.org/methodology-standards.
An update to the Methodology Report will be posted later in 2019.

PATIENT-CENTERED OUTCOMES RESEARCH INSTITUTE

PCORI METHODOLOGY COMMITTEE

Robin Newhouse, PhD, RN, NEA-BC, FAAN *(Chair)*, Distinguished Professor and Dean, Indiana University School of Nursing

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

Naomi Aronson, PhD, Executive Director of Clinical Evaluation, Innovation, and Policy, Blue Cross and Blue Shield Association (BCBSA)

Ethan Basch, MD, MSc, Associate Professor of Medicine, Director, Cancer Outcomes Research Program, University of North Carolina-Chapel Hill

Stephanie Chang, MD, MPH, Director, Evidence-based Practice Centers (EPC) Program, Agency for Healthcare Research and Quality (AHRQ)

David Flum, MD, MPH, Professor, Surgery, Health Services & Pharmacy, Associate Chair for Research, Surgery Director, Surgical Outcomes Research Center, University of Washington

Cindy Girman, DrPH, FISPE, President, CERobs Consulting LLC

Mark Helfand, MD, MS, MPH, Director, Evidence-based Medicine, Professor of Medicine, Professor of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University

Michael S. Lauer, MD, Deputy Director for Extramural Research, National Institutes of Health

David O. Meltzer, MD, PhD, Chief of the Section of Hospital Medicine, Director of the Center for Health and the Social Sciences (CHeSS), Chair of the Committee on Clinical and Translational Science, Associate Professor in the Department of Medicine, Department of Economics and the Harris School of Public Policy Studies, University of Chicago

Brian Mittman, PhD, Research Scientist III, Research & Evaluation, Kaiser Permanente

Sally Morton, PhD, Dean, College of Science, Virginia Tech

Neil Powe, MD, MPH, MBA, Constance B. Wofsy Distinguished Professor and Vice-Chair of Medicine, University of California San Francisco, Chief of Medicine, San Francisco General Hospital

Adam Wilcox, PhD, Professor, Department of Biomedical Informatics & Medical Education, Chief Analytics Officer, University of Washington Medicine

SUGGESTED CITATION

Patient-Centered Outcomes Research Institute (PCORI) Methodology Committee. The PCORI Methodology Report. 2019.

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

TABLE OF CONTENTS

EXECUTIVE SUMMARY

Authorized by the Patient Protection and Affordable Care Act (PPACA) of 2010, the Patient-Centered Outcomes Research Institute (PCORI) 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 healthcare community. PCORI has developed a program of patient-centered outcomes research (PCOR) that meets this goal by emphasizing scientifically rigorous comparative clinical effectiveness research (CER) that examines choices and clinical outcomes that are meaningful to patients and generates evidence that patients and other stakeholders need to improve health and healthcare outcomes.

The PCORI Methodology Committee provides guidance to the institute in advancing this mission and to the research community 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 revised, updated version of the Methodology Report and Methodology Standards adopted by PCORI's Board of Governors in 2017.

This report first addresses the need to take a more systematic approach to prioritizing research topics and determining which research designs can provide information that is both useful and timely to patients, caregivers, clinicians, and other healthcare system stakeholders. PCORI emphasizes the importance of taking a deliberative approach in the translation framework for choosing study designs for specific research questions and considering concerns about the quality of the resulting evidence, appropriate use of scarce research resources, and timeliness of results.

The report then presents the PCORI Methodology Standards. Departures from good research practices are partially responsible for mismatches between the quality and relevance of the information research provides and the information needed to make informed health decisions. The PCORI Methodology Standards help ensure that PCOR studies are designed and conducted to generate the evidence needed to address patients' and clinicians' questions about what works best, for whom, and under what circumstances.

These standards do not represent a complete, comprehensive set of all requirements for high-quality PCOR; rather, they address a group of topics that are likely to contribute to improvements in PCOR quality and value. Specifically, the standards focus on selected methodologies and issues that reflect either areas where there are substantial deficiencies or inconsistencies in how available methods are applied in practice or areas where there is evidence that supports the recommended practices.

Building on the work of the National Academy of Medicine (formerly the Institute of Medicine [2011]), the PCORI Methodology Committee starts 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 should be supported by scientific evidence. When such evidence is unavailable, a standard may be endorsed by reasonable expectation that the standard helps to 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.

The committee then develops the standards by following a systematic process. The committee surveys the range of potential standards, narrows the scope to those it deems most important, solicits feedback through a public comment period, revises the draft standards, and confirms a final set of standards through consensus of its members. In 2018, PCORI added a new standard for data integrity and rigorous analyses as well as a new category of standards for studies of complex interventions.

The current set of PCORI Methodology Standards consists of 56 individual standards in 13 categories. The first five categories of the standards are cross-cutting and relevant to most PCOR studies. Researchers should refer to all of these standards when planning and conducting their projects. These categories are the following:

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

The other eight 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 these data in PCOR studies:

- Data registries
- Data networks as research-facilitating structures

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

- Causal inference methods (**CI-I** applies to all study designs, including randomized trials)
- Adaptive and Bayesian trial designs
- Studies of medical tests
- Systematic reviews
- Research designs using clusters
- Studies of complex interventions

The PCORI Methodology Standards are listed by category in section III of this report. The full text of the standards can also be found in **Appendix A: PCORI Methodology Standards**. PCORI uses the standards in its review of funding applications, monitoring of research awards, and peer review of final research reports submitted by investigators.

This updated set of PCORI Methodology Standards improves the foundation for ensuring best PCOR practices. Given that future advances in research methodology are expected, PCORI has a commitment to continue to evaluate and update the guidance that it provides to the research community.

INTRODUCTION

Authorized by the Patient Protection and Affordable Care Act of 2010, the Patient-Centered Outcomes Research Institute (PCORI) was established to help people make informed healthcare decisions and improve healthcare delivery and outcomes by producing comparative clinical effectiveness research (CER) that is guided by patients, caregivers, and the broader healthcare community. According to the National Academy of Medicine (formerly the Institute of Medicine), comparative clinical effectiveness research (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" (Institute of Medicine 2009). PCORI has developed a program of patient-centered outcomes research (PCOR) that meets this goal by emphasizing scientifically rigorous research that examines choices and clinical outcomes that are meaningful to patients and generates evidence that patients and other stakeholders need to improve health and healthcare outcomes.

The federal legislation 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 was to develop 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 was to create a translation table that would provide guidance to "determine research methods that are most likely to address each specific research question." PCORI completed its initial work on these requirements in 2012 and released the first edition of this report in 2013.

PCORI developed an initial set of methodology standards designed to improve the conduct of patient-centered CER (PCORI Methodology Committee 2013) and updated the standards in 2017 (PCORI Methodology Committee 2017). In 2018, PCORI added a new standard for data integrity and rigorous analyses as well as a new category of standards for studies of complex interventions. These new standards were posted for public comment in the last half of 2017. The new standards are listed in section III of this report, which provides the rationale for each set of standards and additional discussion about the methodological issues that the standards are intended to address.

This report also addresses the need to take a more systematic approach to prioritizing research topics and determining which research designs are most appropriate for generating the strong, high-quality findings needed to address clinical evidence gaps. Section II outlines key considerations and decision points in the research process that are critical to ensuring that PCOR studies provide information that is both useful and timely to patients, caregivers, clinicians, and other healthcare system stakeholders.

To illustrate the importance of the issues addressed in this report, we have included four sets of stories and examples collected in 2013, 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 the importance of using appropriate methods to ensure the validity, trustworthiness, and usefulness of findings generated by PCOR.

PATIENT VOICES

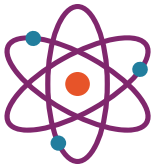
Focus on patients who share their experiences in navigating choices and weighing options.

RESEARCH STORIES

Focus on published research studies that demonstrate the importance of good methodology for producing valid and useful research results.

CER WINS

Focus on comparative effectiveness research (CER) that led to important changes in clinical practice and patient care.



RESEARCH IN PRACTICE

Focuses on the value and challenges of implementing CER studies.

SECTION I: PATIENT-CENTERED OUTCOMES RESEARCH

The availability of multiple options for prevention, diagnosis, and treatment in health care presents a significant challenge to patients and clinicians trying to make informed health decisions. Deciding between healthcare options requires not only an understanding of how to balance the benefits and risks of each treatment option but also an understanding of how each option might apply differently to individual patients, given their unique personal characteristics. The information needed to make these decisions most often comes from clinical research.

A program of clinical research should provide high-quality, relevant, and useful health-related evidence for decision makers, especially patients, caregivers, and healthcare providers. Patient-centered outcomes research focuses on providing information that can help patients address questions such as the following:

- Given my personal characteristics, conditions, and preferences, what should I expect will happen to me?
- What are my options, and what are their potential benefits and harms?
- What can I do to improve the outcomes that are most important to me?
- 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 effectiveness research can be found in **CER Wins: Two Studies of Improving Care in Hospitals**.)

Frequently, however, a gap exists between the information that people need for informed health decisions and the information available from research. This gap sometimes results from how research questions are selected, how studies are designed, and how results are disseminated. Researchers often choose questions and outcomes that they consider interesting and important, but sometimes these are not the questions and outcomes that are most relevant to people who need information. Researchers may be less inclined to focus on outcomes that are difficult to obtain, 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**.)

Often, research is conducted with individuals who represent only a limited range of characteristics, such as age, sex, race, and complexity of conditions. Some research also may be restricted to treatment in sophisticated research centers rather than typical community settings. Practical reasons may influence these choices: it takes a much larger study to account for differences among patients, and the bigger the study, the more the research costs. Conducting research in multiple settings or community settings where research is less common takes more work. Sometimes researchers want to include a broader range of patients and settings but are unable to do so because they have trouble either recruiting study participants who represent the full spectrum of patients or managing the logistics of multiple sites. (To learn about two trials that used broader inclusion criteria, see **CER Wins: The Value of Including a Wide Variety of Patients and Settings in Studies**.)

Moreover, comprehensive reviews of research have shown that many studies address questions that have already been answered, fail to address questions that are widely known to be important, or use study designs that render the results useless for decision makers (Chalmers and Glasziou 2009; Macleod et al. 2014). Failure to conduct fair "head-to-head" comparisons of alternative treatments (Evans et al. 2011), employ appropriate methods (Yordanov et al. 2015), and ensure full publication of study results (Glasziou et al. 2014), including negative and null findings, represent significant sources of "avoidable waste" in research and contribute to the persistence of evidence gaps (Chalmers and Glasziou 2009).

CER WINS

Two Studies Using Hospitals to Improve Care

Comparative clinical 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 seek to determine which strategies are most effective.

What Strategies Help Hospitals Avoid Infections?
Too often, patients get infections while in the hospital, and such hospital-acquired infections 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. Other states exhibited similar rates. But such infections are often preventable.

A major source of the infections are thin tubes, called central line catheters, inserted into large veins. In the Keystone Intensive Care Unit (ICU) project, most Michigan hospitals participated in a large, prospective, observational study that examined a new process to prevent 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 patient safety everyone's job. The team compared Michigan hospitals that made the changes with hospitals in nearby states that did not. After two years, among patients 65 years or older, no catheter-associated infections occurred 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 to conduct this procedure, so patients who need angioplasty are often transferred to hospitals that offer it.

Randomized controlled trials have compared patients who were moved and received angioplasty with those treated in other ways at the original hospital. When no delays occurred, 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 patient data registries 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 (drug) therapy at the first hospital. The results demonstrated that delays in 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 two hours (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.

PCORI is committed to addressing these challenges and supporting high-quality PCOR. The PCORI Methodology Standards have been developed to address specific criticisms and weaknesses of clinical research. These standards establish expectations about the characteristics of high-quality PCOR, specifying a set of requirements for scientifically valid, transparent, and reproducible research. Consistent with the objectives of these standards, PCORI is committed to the principles of open science, which is broadly defined as efforts to increase meaningful public and professional access to the results and data from research. Improving transparency of, access to, and utility of data from clinical research can facilitate the reproduction of original analyses (allowing other researchers to verify the findings) as well as the conduct of additional analyses (improving research efficiency and the responsible use of limited research resources). PCORI believes that for evidence to be useful, it must be relevant and readily available to the people who are making decisions (see **Research in Practice: Chest Pain Choices**), and PCORI supports efforts to improve public access to study reports for all relevant stakeholders.

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 much 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 shown 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 looked only at intermediate outcomes (abnormal heart rhythms). The trial led to an immediate and lasting change in treatment for patients who had previously had a heart attack.

The Value of Including a Variety of Patients and Settings in Studies

Some randomized trials of medical treatments use strict eligibility criteria to select people who are similar to one another: all of the participants 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 the First International Study of Infarct Survival (ISIS-1), enrolled 16,000 people in 14 countries. Each person had experienced symptoms of a heart attack and had gone to a hospital. 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 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.

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 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.

SECTION II: IDENTIFYING AND ADDRESSING EVIDENCE GAPS IN PATIENT-CENTERED OUTCOMES RESEARCH

Establishing a specific research agenda is one of PCORI's core duties. 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 should be directed toward providing the answers patients, clinicians, and other stakeholders need for health decisions.

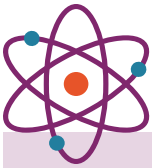
Identifying and Prioritizing Research Questions

PCORI's Board of Governors is charged with identifying research priorities, developing a research agenda based on those priorities, and funding studies that align with those priorities. PCORI's national research priorities are informed by the following considerations:

- 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
- Relevance to patients and clinicians in making informed health decisions

PCORI is obligated to spend its resources effectively and efficiently. When more than one acceptable research approach is available, the advantages and disadvantages of alternative study designs should be considered, including the potential value and timeliness of the likely research results. Techniques such as value-of-information analysis—a statistical method for estimating the average improvement in outcomes that may be expected by obtaining additional information (Claxton and Sculpher 2006; Meltzer et al. 2011)—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**). However, such tools cannot replace reasoned judgment and transparent discussions between decision makers and relevant stakeholders in determining the level of evidence needed to support informed health decisions and how best to generate it.

PCORI must consider a sufficient number and range of topics before it selects topics for research funding. Including patients and other stakeholders can help to better align new research topics with the information needs of patients, clinicians, and other healthcare stakeholders (Sheridan et al. 2017). Empirical evaluations of engagement in research increasingly suggest that the involvement of patients and other stakeholders can improve the relevance of research questions and usefulness of results for health decision making (Dudley et al. 2015; Esmail, Moore, and Rein 2015; Forsythe et al. 2016). PCORI is therefore exploring novel and existing approaches to obtaining patient and other stakeholder input in research topic generation (see **Research in Practice: PCORI Prioritization Pilot**). PCORI is also systematically evaluating the impact of patient and other stakeholder engagement on the research it funds to identify best practices for engagement in PCOR studies (Frank, Basch, and Selby 2014).



RESEARCH IN PRACTICE

Analyzing the Value of Information

Originally published in 2013

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 VOI. 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 II 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?
Carlson: In our study, we asked 13 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.

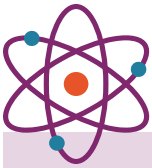
Systematic Reviews

Research funders have an ethical obligation to avoid involving patients in unnecessary studies. A study is needed if it addresses an important question that has not been answered by previous research—namely, if it addresses an "evidence gap." Systematic reviews, which critique and synthesize the existing literature, can identify gaps in knowledge that underlie uncertainty among patients and clinicians. Systematic reviews can also highlight key questions that have not been answered by prior studies. Identifying gaps in the existing literature and deficiencies in completed studies can reduce investments in research that are unlikely to help answer important questions.

Peer and Stakeholder Review of Research Proposals

Despite its central role in scientific discourse and decision making, peer review of research proposals has had little attention as a subject of research; most peer-review practices are maintained by convention (Kotchen and Spellecy 2012). At PCORI, research proposals are reviewed by scientists, patients, and other healthcare stakeholders. PCORI has chosen to involve patients and other stakeholders in the review process because of the central importance of patient centeredness (Fleurence et al. 2014; see **Patient Voices: PCORI Reviewers**).

To protect the integrity and independence of the review process, PCORI has sought to adhere to strict standards for avoiding conflicts of interest. Research proposals are also assessed for adherence to PCORI's Methodology Standards to ensure that the research selected for funding is designed to generate the high-quality and relevant evidence needed to inform health decisions.



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, former director of PCORI's CER Methods and Infrastructure program, stressed 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 multistakeholder 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 and 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.

Designing Research to Address Evidence Gaps

After evidence gaps have been identified and prioritized, PCOR studies must be designed to generate the evidence needed to close these gaps and provide the information necessary to make informed health decisions. The quality and relevance of evidence generated by a study depends not only on the study's design but also on the choice of data source(s) and analytical methods. Regardless of the choices made, there will always be limitations in the design, implementation, and analysis of clinical research. The key is to ensure that these limitations are recognized and that steps are taken to minimize the risk that a study will produce biased results with serious consequences for patients (e.g., overestimating the benefits of treatments, underestimating the harms).

PCORI's authorizing legislation directs the organization to develop a translation table to help its Board of Governors understand the study design(s) and methods that are most likely to address a specific comparative clinical effectiveness research question. Although this directive implies a one-to-one relationship between a research question and choice of study design, it is widely accepted that most research questions can be answered using many kinds of designs. The choice of study design and method is multifaceted, complex, and based on several factors; there is no formula that can be applied to all situations in PCOR.

Therefore, PCORI has outlined a translation framework that reflects a deliberative process for guiding the choice of study designs for specific research questions and the key elements that need to be considered to ensure the quality of the resulting evidence, appropriate use of scarce research resources, and timeliness of results (see appendix C). The framework is not intended to be directed toward a specific choice of design and methods but toward deliberation about the options and trade-offs at each decision point in the research process and how best to accomplish the research objectives. Methodological expertise is needed in these discussions to weigh the options, priorities, and available resources when choosing a study design.

The research process begins by generating patient-centered research questions. The components (often abbreviated as

PICOTS) of a well-formulated research question include the following (see, e.g., Richardson et al. 1995):

- **P**opulation of patients/research participants and relevant subgroups of patients
- **I**ntervention(s) relevant to patients in the target population
- **C**omparator intervention(s) relevant to patients in the target population
- **O**utcomes that are meaningful to patients in the target population
- **T**iming of when outcomes are assessed and length of follow-up
- **S**ettings in which the intervention is delivered, including the healthcare providers

Multiple perspectives—including those of patients, clinicians, researchers, policy makers, and other stakeholders—may shape the research question. Regardless of the process used to generate the research question, the decision that the study is meant to inform must be clearly defined, and a systematic review (or other critical appraisal) of prior studies should be undertaken.

The choice of research question should (at least initially) be kept distinct from discussions about the methodology. The available approaches to study design and analysis represent the potential options for addressing a selected research question, and problems can 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, although these constraints may inform decisions about the extent to which a particular research question can be adequately addressed by a new study.

Once the research question has been formulated, the potential design options can be considered. The choice between a randomized or observational design is based on many factors, including timeliness, representativeness, validity of findings, data quality, and the ability to identify subgroup effects. Such study characteristics (see **Examples of Study Characteristics**) influence the usefulness of the results for decision making. There is usually more than one acceptable choice. For example, to obtain results sooner and/or enhance external validity, an observational study that uses secondary data (information from previously collected data) could be considered; however, this design would likely have more threats to internal validity than would an experimental study that uses randomization. However, the experimental study could fail to address the research question if it is not representative of care (and the decisions faced by patients and clinicians) outside the controlled research environment. Logistical issues can also be as challenging as scientific ones. For example, if only a limited number of patients with a specific condition are available to study, then sampling and data collection strategies using existing healthcare data sources might be needed to successfully conduct the study.

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 the following: 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.

"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 health care … 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 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

Advances in research methodology should also be considered. Over the past 20 years, choice of study design has been debated intensely in scientific venues. Some assert that randomized designs are more relevant than observational studies to decision makers, but well-designed observational studies have also demonstrated value individually or as a complement to randomized designs, helping to determine under what circumstances and to which patients the findings from randomized controlled trials (RCTs) apply. Observational studies also may uncover rare events (often harms) that were not observed in RCTs. The potential for observational studies to support causal inferences is much stronger than it has been in the past (Institute of Medicine 2012, 2013).

The selection of either a randomized or observational study is only a starting point, however. The choice of data source(s) and analytical methods also affects the strength and quality of evidence generated by a study (Institute of Medicine 2012). Important considerations include, for example, whether the nature of the study question requires that specific information be newly collected, or whether information from previously collected data will suffice. If data have been previously collected, several factors should be considered, including availability of clinical detail, data completeness, access to the data, confidentiality, and ability to link multiple data sources. Analytical methods should be selected to address issues of bias and confounding that could invalidate estimates of an intervention's benefits and risks.

EXAMPLES OF STUDY CHARACTERISTICS

Intrinsic Study Characteristics	Extrinsic Study Characteristics
• Internal validity: the extent to which effects are caused by the intervention or exposure	• Timeliness: rapidly changing technology, policy, or public health needs
• External validity: generalizability or applicability to non-study settings and populations	• Logistical constraints: feasibility of collecting information from participants, number of participants available, study complexity
• 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	

A core tenet of PCOR is that the perspectives of patients and other stakeholders can inform scientific reasoning about the research hypothesis and research question(s), elements of study design and conduct, and outcome selection and measurement; these perspectives also help to ensure that studies provide answers to real-life "decisional dilemmas" and improve health outcomes. Regardless of the source, input from stakeholders must be examined for its scientific validity and potential to strengthen the research. Therefore, thoughtful deliberation among researchers, patients, clinicians, and other stakeholders is needed to determine which research designs and methods will provide valid and useful information to fill today's clinical evidence gaps.

SECTION III: PCORI METHODOLOGY STANDARDS

Introduction

Because patient-centered outcomes research can include a variety of research designs and specific techniques, PCORI's Methodology Standards do not attempt to address all possible issues in clinical research. Rather, the topics for the standards were chosen to reflect areas where there were either (1) substantial deficiencies or inconsistencies in how available methods were applied in practice, despite specialized knowledge about how best to conduct research; or (2) threats to the validity of research results that diminished the value and potential use of those results (Helfand et al. 2011; Lohr 2007; Schneeweiss, Seeger, and Smith 2012).

Background

After following a structured process to obtain input from scientific experts and solicit public comments, PCORI's Board of Governors endorsed an initial set of standards that was released to the public in December 2012. Details on the standards development process were provided in the first edition of this report (PCORI Methodology Committee 2013).

Workgroups of Methodology Committee members and PCORI staff periodically review PCORI's Methodology Standards to update existing standards and develop additional standards. Each workgroup evaluates the methodological literature, and outside consultants are engaged as needed. Through a consensus process, each workgroup proposes new or updated standards, which are reviewed and revised by the full committee, posted for public comment, and finalized following additional revisions based on public comments.

Using this systematic approach, the Methodology Committee developed a new set of standards for studies of complex interventions as well as a new standard to improve data management throughout the research process. The new standards were posted on the PCORI website, and public comments were solicited between October and December 2017. Following the public comment period, the Methodology Committee made further revisions to the revised standards. The PCORI Board of Governors adopted these standards in April 2018. (The table in appendix B summarizes the response to public comments.) The current PCORI Methodology Standards, which are discussed in this report, consist of 56 individual standards in 13 categories (see **Appendix A: PCORI Methodology Standards**.)

Overall Rationale

PCORI's efforts to establish methodological standards for PCOR are a logical extension of other efforts to improve research methodology. 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; US Food and Drug Administration 2010a, 2010b). These standards, articulated in formal "guidance documents," helped to create a level playing field for companies that design 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, PCORI's Methodology 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 standards that address only the reporting of results after studies are completed. Reporting standards for different study designs are currently housed at the Equator network website, 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 National Academy of Medicine (NAM), formerly the Institute of Medicine, 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" (Institute of Medicine 2008). In 2011, NAM published standards for conducting systematic reviews (Institute of Medicine 2011). The PCORI Methodology Standards expand this effort by formulating criteria for comparative clinical effectiveness research such as randomized trials, observational studies, and studies of medical tests.

As a group, the PCORI Methodology Standards offer an approach to ensuring that PCOR studies are designed and conducted to generate the evidence needed to address patients' and clinicians' questions 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 and provide a common set of expectations about the characteristics of high-quality PCOR.

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

- Formulating research questions
- Patient centeredness
- Data integrity and rigorous analyses
- Preventing and handling missing data
- Heterogeneity of treatment effects

The other eight categories of standards apply to particular study designs and methods. Two of the categories provide guidance on developing specific types of data and using these data in PCOR studies:

- Data registries
- Data networks as research-facilitating structures

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

- Causal inference methods (**CI-I** applies to all study designs, including randomized trials)
- Adaptive and Bayesian trial designs
- Studies of medical tests
- Systematic reviews
- Research designs using clusters
- Studies of complex interventions

These standards should be considered minimal standards, meaning that they are necessary for sound science but should not discourage the use of more sophisticated approaches and/or inhibit further evolution of methods. Some standards are designed to promote transparency: how to properly communicate, both in study protocols and in published reports, exactly what was planned and what was done. All of the standards are based on current scientific knowledge; some standards are based on theoretical work and/or simulations because evidence from empirical studies was not available.

The following sections present the standards, grouped by category. Each section begins with the full text of all standards in that category, followed by a brief summary of the rationale for the standards, key definitions, and additional discussion about the methodological issues. References to the applicable standard are included in parentheses—for example, (**RQ-1**).

1: STANDARDS FOR FORMULATING RESEARCH QUESTIONS

RQ-1: Identify gaps in evidence.

Gaps in the evidence identified in current systematic reviews should be used to support the need for a proposed study. If a systematic review is not available, one should be performed using accepted standards in the field (see **SR-1**), or a strong rationale should be presented for proceeding without a systematic review. If the proposed evidence gap is not based on a systematic review, the methods used to review the literature should be explained and justified.

RQ-2: Develop a formal study protocol.

Researchers should develop a formal protocol that provides the plan for conducting the research. The protocol should specify the research objectives, study design, exposures and outcomes, and analytical methods in sufficient detail to support appropriate interpretation and reporting of results. Protocols should be submitted to the appropriate registry (e.g., clinicaltrials.gov), and all amendments and modifications (e.g., changes in analytic strategy, changes in outcomes) should be documented.

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(s) for whom the health decision is pertinent, and (3) how study results will inform the health decision.

RQ-4: Identify and assess participant subgroups.

When designing studies, researchers should identify participant subgroups, explain why they are of interest, and specify whether subgroups will be used to test a hypothesis or to conduct an exploratory analysis, preferably based on prior data. A study should have adequate precision and power if conclusions specific to these subgroups will be reported.

RQ-5: Select appropriate interventions and comparators.

The interventions and comparators should correspond to the actual healthcare options for patients, providers, and caregivers who would face the healthcare decision. The decision should be of critical importance to the relevant decision makers, and one for which there is a compelling need for additional evidence about the benefits and harms associated with the different options. Researchers should fully describe what the comparators are and why they were selected, describing how the chosen comparators represent appropriate interventions in the context of the relevant causal model (**CI-1**), reduce the potential for biases, and allow direct comparisons. Generally, usual care or nonuse comparator groups should be avoided unless they represent legitimate and coherent clinical options.

RQ-6: Measure outcomes that people who represent the population of interest notice and care about.

Identify and include outcomes that the population of interest notices and cares about (e.g., survival, functioning, 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 that reflect both beneficial and harmful effects, based on input from patient informants and people representative of the population of interest.

Rationale for These Standards

A primary objective of PCOR is to enable patients and those who care for them to make better informed decisions by generating strong and high-quality evidence about the risks and benefits of their available healthcare options. As with other approaches to clinical research, PCOR involves four broad phases, or categories, of scientific activities:

- Formulation of the research question ("What should we study?")
- Selection of the study approach ("What study design[s] should we use?")
- Execution of the study ("How do we conduct, govern, and analyze the study?")
- Dissemination and implementation of findings ("How do we enable people to apply the study results?")

Many of the PCORI Methodology Standards focus on the early phases of research, because 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 healthcare decisions, that recruitment strategies will achieve the participant numbers required for scientific rigor, and that the protocol makes clear how the research will accomplish its objectives. These (and other) standards specify what to include in research protocols as a means of improving the quality of the study and the transparency of the research process. Higher quality and more transparent research should result in a better understanding of the applicability of study results to specific patients and situations.

Getting the questions right ("What should we study?") is an important starting point. The Standards for Formulating Research Questions provide guidance in determining whether additional research is needed to support informed health decisions and how to ensure that studies are designed to generate the necessary information.

The need for a new study must be rigorously justified. To make optimal use of resources available for research, study questions should not be redundant or irrelevant to healthcare practice and decisions. Proposed research projects should address gaps in knowledge about treatments or services, including gaps in understanding what works in populations that differ from those that have been studied (e.g., studies in different age or socioeconomic groups). Research imposes risk on participants (even secondary analyses of data can present risks, such as the disclosure of sensitive information), and the imposition of these risks cannot be justified if the research will not provide evidence to improve health decisions.

Careful, thorough consideration of previous and continuing studies can help prevent wasted investments in research (Ioannidis et al. 2014). Systematic reviews play a critical role in the justification of research, supporting a structured approach to assessing not just whether there is a lack of evidence but whether that lack of evidence demonstrably hinders the ability of patients, caregivers, and providers to make an informed decision about their health and health care (Chalmers et al. 2014). If a systematic review is not available—and if conducting one may not be useful or the best use of resources—researchers should describe and justify the approach employed to identify the evidence gap, including any departures from relevant standards for conducting and reporting systematic reviews (see Standards for Systematic Reviews) (**RQ-1**).

Once the need for new research is established, a formal study protocol should be developed that provides a comprehensive plan for the design, conduct, and analysis of the study (**RQ-2**). Formal protocols make the study intentions clear to all users, provide the information needed to evaluate the quality and applicability of the research, and help to ensure that spurious results are not reached because of multiple post hoc analyses.

The research question and study protocol should clearly describe the following components (often abbreviated as PICOTS), which are captured in **RQ-3** through **RQ-6**:

- **P**opulation of patients/research participants and relevant subgroups of patients
- **I**ntervention(s) relevant to patients in the target population
- **C**omparator intervention(s) relevant to patients in the target population
- **O**utcomes that are meaningful to patients in the target population
- **T**iming of when outcomes are assessed and length of follow-up
- **S**ettings in which the intervention is delivered, including those of the healthcare providers

Describing who is included (and excluded) in the study population is essential for understanding to which patients and in what circumstances the results will apply as well as for ensuring the reproducibility of study findings (**RQ-3**). Many studies also aim to determine how the treatments being compared affect significant subgroups of the population (**RQ-4**) or use subgroup analysis to generate ideas for future research. However, subgroup analyses may not always be appropriate, depending on the research question, size of the subgroups, and available evidence (see the section on Standards for Heterogeneity of Treatment Effects for additional discussion). The selection of comparators (**RQ-5**) and outcomes (**RQ-6**) should be justified with respect to the specific evidence gap and health decision that the study is designed to address (see the Standards Associated with Patient Centeredness for additional discussion related to **RQ-6**). Notably, the choice of outcome measures—not just the choice of outcomes—can affect the interpretability, validity, and relevance of results (Velentgas, Dreyer, and Wu 2013); explicit justification should be provided for decisions about how to operationalize and measure the outcomes of interest.

2: STANDARDS ASSOCIATED WITH PATIENT CENTEREDNESS

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

Include individuals affected by the condition and, as relevant, their surrogates and/or caregivers. Other relevant stakeholders may include, but are not limited to, clinicians, purchasers, payers, industry, hospitals, health systems, policy makers, and training institutions. These stakeholders may be end users of the research or be involved in healthcare decision making.

As applicable, researchers should describe how stakeholders will be identified, recruited, and retained and the research processes in which they will be engaged. Researchers should provide a justification in proposals and study reports if stakeholder engagement is not appropriate in any of these processes.

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 the following:

- The plan to ensure representativeness of participants
- How participants are identified, selected, recruited, enrolled, and retained in the study to reduce or address the potential impact of selection bias
- Efforts employed to maximize adherence to agreed-on enrollment practices
- 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 (e.g., individuals historically underrepresented in healthcare research, such as those with multiple disease conditions, low literacy, low socioeconomic status, or poor healthcare access; racial and ethnic minority groups; people living in rural areas), then specify plans to address population-specific 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 for outcomes of interest.

To measure outcomes of interest identified as patient centered and relevant to decision makers (see **RQ-6**) and for which patients or people at risk of a condition are the best source of information, the study should employ patient-reported outcome (PRO) measures and/or standardized questionnaires with appropriate measurement characteristics for the population being studied. When selecting PRO measures for inclusion in a study, researchers, in collaboration with patient and other stakeholder partners, should consider (1) the concept(s) underlying each PRO measure (e.g., symptom, 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 its 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 as well as the translation process if the measure is to be used in multiple languages. If these measurement properties are not known, a plan to establish the properties must be provided. Caregiver reports may be appropriate if the patient cannot self-report the outcomes of interest.

PC-4: Support the dissemination and implementation of study results.

All study results must be made publicly available. To ensure study objectives and results are understandable and actionable by as many people as possible, they should be presented in lay language summaries. For study results that are appropriate for dissemination and implementation, involve patients and other relevant stakeholders in (1) planning for dissemination from the start of the research study, (2) creating a dissemination plan for the study that indicates clinical implications, (3) working with patients or organizations to report results in a manner understandable to and usable by each target audience, and (4) identifying successful strategies for the adoption and distribution of study findings to targeted patient and clinical audiences.

PATIENT VOICES

Lucinda Shore

Originally published in 2013

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 a 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," she 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 that 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."

Rationale for These Standards

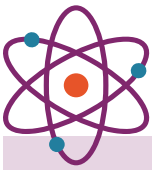
The purpose of PCOR is to help people make informed healthcare decisions. To do this, PCORI must direct research toward addressing questions that are important to patients, measure outcomes that are noticeable and meaningful to them, and produce results that help them assess the value of healthcare options, given their personal circumstances, conditions, and preferences. The standards in this group are designed to improve the quality and relevance of PCOR findings by supporting effective engagement of patients and other stakeholders and by explicitly incorporating patient needs, values, and preferences.

In addition to supporting meaningful and systematic approaches for engaging patients and other stakeholders throughout the research process, these standards should facilitate improved understanding of how such engagement affects study design and outcomes through improved reporting of patient-centered research processes. The increased emphasis on patient and other stakeholder engagement in the research process reflects not only a commitment to important values of social justice and democratic participation (Domecq et al. 2014; Esmail, Moore, and Rein 2015) but also the hypothesis that such engagement will improve the quality and relevance of the research (Carman et al. 2013). Although the empirical evidence underlying early guidelines and recommendations for the inclusion of patients and other stakeholders in research was limited and varied considerably in quality (Gagnon et al. 2011; Staniszewska et al. 2011), systematic efforts to evaluate the impact of patient and other stakeholder engagement on the quality of research are underway (Frank et al. 2015). Early findings suggest an effect of engagement on study design (including selection of comparators and outcomes), recruitment, and retention (Dudley et al. 2015; Forsythe et al. 2016).

To ensure patient centeredness, researchers should describe and report their plans for engaging those who represent the population of interest and other relevant stakeholders (i.e., how they will partner with them in appropriate phases of research) (**PC-1**). 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**). This engagement can include, for example, getting patients and other stakeholders to help identify topics and formulate research questions; identify a study population and choose interventions, comparators, and outcomes; develop and implement optimal strategies to recruit and retain study participants; conduct and monitor the study (including data collection activities); analyze data and interpret the findings; and disseminate the results (Frank, Basch, and Selby 2014; Mullins, Abdulhalim, and Lavallee 2012).

Researchers should ensure that study participants are 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 this standard, PCORI evaluated specific strategies for involving people who have been historically underrepresented in research or who may be difficult to reach (Mullins, Barnet, et al. 2012). Participant recruitment and retention in general and minority recruitment and retention in particular are known to be significantly subpar in clinical research.

When patients and other stakeholders are engaged as research partners, they play a critical role in addressing the



RESEARCH IN PRACTICE

Pamela Williams

Originally published in 2013

Millions of Americans with rare diseases not only often deal with misdiagnoses, diagnostic delays, and a frustrating search for treatments, but they also may experience social and psychological problems that the healthcare system doesn't recognize. Pamela Holtzclaw Williams, PhD, JD, RN, wants to change that. Williams, a University of Arkansas researcher, was awarded a PCORI contract to use feedback from patients with the rare disease alpha-1 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. Decisions faced by those with rare genetic illnesses include the following: Who gets tested in the family? Who should 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 benefit 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

Originally published in 2013

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 to 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 composed of an oncologist, a surgical oncologist, a plastic surgeon, a radiation oncologist, certified nurse practitioners, and nurses. "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 had reservations about her treatment and expected outcomes, and looked to Simmons to help communicate them. 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 well-being. 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."

aforementioned challenges. Robust engagement approaches can strengthen the recruitment and retention of study participants and ensure the successful conduct of research. Examples of such approaches include community advocate training, community and stakeholder advisory boards, and collaborations with outside groups (e.g., healthcare providers, service delivery sites, community-based organizations) to promote referrals and inquiry.

Patient centeredness in research also requires the identification, measurement, and evaluation of outcomes that are meaningful to patients (see also **RQ-6**). Researchers and patient and stakeholder partners should identify the outcomes of interest and select the appropriate outcome measures. In cases where patients or people at risk of a condition are the best source of information about a particular outcome of interest, studies should employ PRO measures and/or standardized questionnaires with appropriate measurement characteristics for the population being studied (**PC-3**). PROs are health data reported by the patient "without amendment or interpretation of the patient's report by a clinician or anyone else" and measured by self-report or interview (American Institutes for Research 2016; US Food and Drug Administration 2015). PROs are particularly important in assessing the effects of an intervention on symptoms or other outcomes (e.g., pain) that are only directly known by the individual patient, but they can be also be used to assess patient perspectives on outcomes (e.g., functioning) that may be observable to others (US Food and Drug Administration 2015).

The standards do allow for the development and evaluation of new PRO measures, when justified, to measure outcomes that are important to patients (see **Patient Voices: Juli** and **Patient Voices: A Woman with Fibromyalgia**). There also may be specific circumstances (e.g., studies of infants or people with severe cognitive impairment) in which the most suitable outcome measure(s) would be based on the reports of caregivers or assessments of observable behaviors (e.g., facial expressions). In cases where patients cannot provide direct reports, caregiver reports of observable signs or events are preferred over reports of symptoms (e.g., pain) that require interpretation by the observer (US Food and Drug

Administration 2015). Other sources of information, including clinician reports and administrative data (e.g., length of hospital stay), can also provide data on outcomes that are meaningful to patients and other end users of the research.

To conclude the patient-centered research process, dissemination of the study's findings should integrate the new results with related work and underscore meaningful clinical and policy implications from the perspective of patients and other stakeholders. Although, in rare cases, dissemination of research findings beyond traditional mechanisms of scientific publications and presentations may be outside the scope of an individual research project, researchers should work with patients and other stakeholders to support efforts for effective dissemination and implementation of results (**PC-4**). 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. Any successful implementation strategy must also identify and mitigate barriers to the adoption of clinical strategies that are informed by the study's findings. Researchers should work with their stakeholders to identify such barriers and to develop and refine dissemination plans prior to study completion.

PATIENT VOICES

A Woman with Fibromyalgia

Originally published in 2013

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 rheumatologist's 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 me to rate, 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; rather, it was because I wanted to be strong and not give in to the pain. I said to myself, "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 where 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.

3: STANDARDS FOR DATA INTEGRITY AND RIGOROUS ANALYSES

IR-1: A priori, specify plans for quantitative data analysis that correspond to major aims.
Before analysis is undertaken, researchers should describe the analytic approaches that will be used to address the major research aims. These include definitions of key exposures, outcomes, and co-variates. As applicable, study protocols should identify patient subgroups of interest, plans (if any) for how new subgroups of interest will be identified, and how analysis plans may be adapted based on changing needs and scientific advances. Researchers should also specify plans for handling missing data and assessing underlying assumptions, operational definitions, and the robustness of their findings (e.g., sensitivity analyses).

IR-2: Assess data source adequacy.
In selecting data sources and planning for data collection, researchers should ensure the robust capture of exposures or interventions, outcomes, and relevant covariates. Measurement properties of exposures and outcomes should be considered, and properties of important covariates should be taken into account when statistically adjusting for covariates or confounding factors.

IR-3: Describe data linkage plans, if applicable.
For studies that link patient data from two or more sources (including registries, data networks, and others), describe (1) the data sources and/or the linked data set in terms of 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(s).

IR-4: Document validated scales and tests.
Studies should include documentation of the names of the scales and tests selected, reference(s), characteristics of the scale, and psychometric properties.

IR-5: 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 lists 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), STROBE (for observational studies), and SRQR and/or COREQ (for studies that use qualitative research). Researchers should register their studies with the appropriate registry (e.g., clinicaltrials.gov for clinical studies or observational outcomes studies) and provide complete and accurate responses to the information requested (e.g., enter the required and optional data elements for clinicaltrials.gov).

IR-6: Masking should be used when feasible.
Masking (also known as blinding) of research staff should be implemented, especially in situations for which study participant and investigator masking are not feasible. When masking is not feasible, the impact of the lack of masking on the results should be discussed.

IR-7: In the study protocol, specify a data management plan that addresses, at a minimum, the following elements: collecting data, organizing data, handling data, describing data, preserving data, and sharing data.
Data management is a critical phase in clinical research that contributes to the generation of high-quality, reliable, and statistically sound data from clinical trials and observational studies. The underlying motivation for good data management practice is to ensure that the data are accessible, sustainable, and reproducible, both for future investigators and for the original research team. This standard applies to both the quantitative and the qualitative data collected in a study.

A data management plan (DMP) is a document that describes what data will be generated by a research study, how the data will be managed and stored, who will have access to the data, what documentation and metadata will be created with the data, how the data will be preserved, and how the data will be shared in support of future scientific inquiries.

DMPs are distinct from statistical analysis plans, which describe the planned statistical analyses associated with the study (e.g., statistical tests to be used to analyze the data, how missing data will be accounted for in the analysis).

To ensure quality control, the study investigators should self-monitor their data management procedures. This includes conducting checks to ensure manually entered subject numbers conform to study-defined site/subject number format rules and conducting real-time review of data to verify their accuracy and validity.

DMPs should include language that, at a minimum, addresses each of the following considerations:

- Collecting data: Based on the hypotheses and sampling plan, describe what data will be generated and how they will be collected. Provide descriptive documentation of the data collection rationale and methods, and any relevant contextual information.
- Organizing data: Decide and document how data will be organized within a file, what file formats will be used, and what types of data products will be generated.
- Handling data: Describe and document who is responsible for managing the data, how version control will be managed, what the data handling rules are, what the method and frequency for backing up the data will be, and how confidentiality and personal privacy will be protected.
- Describing data: Describe how a data dictionary and metadata record will be produced (i.e., metadata standard and tools that will be used).
- Storing and preserving data: Implement a data storage and preservation plan that ensures that both the raw data and the analytic files can be recovered in the event of file loss. Document the data storage and preservation plan, including the approach to data recovery (e.g., routinely storing data in different locations).
- Maintaining data: Develop a plan to maintain the data in a data repository.
- Sharing data: Develop a plan to share data with the project team, with other collaborators, and with the broader scientific community.

Consistent with the Guideline for Good Clinical Practice, the investigator/institution should maintain adequate and accurate source documents, including the DMP. The DMP should be attributable, contemporaneous, original, accurate, and complete. Changes to the DMP should be traceable, should not obscure the original entry, and should be explained if necessary (e.g., via an audit trail).

Rationale for These Standards

The standards that address data integrity and analysis methods build on best practices in clinical research and add to several other categories of standards (including the Standards for Formulating Research Questions) by requiring documentation of key decisions about data collection and measurement as well as the assumptions made in the analyses. These standards emphasize prospective specification of the research design elements related to data and analyses to determine whether data are likely to be adequate to address the proposed research questions before the research begins. These standards apply to research that employs quantitative, qualitative, and/or mixed-method approaches and address whether the research uses existing data, involves primary data collection, or combines data from multiple sources.

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 but should also be considered when planning primary data collection. Assessing data adequacy involves determining whether data on important outcomes as well as other factors that could affect results (e.g., mitigating factors, confounding factors) are available and valid (**IR-1** and **IR-2**).

To allow users of the research findings to evaluate whether the study produced reliable results and the extent to which results generalize to other settings and populations, researchers must describe the decisions they made about the design and conduct of analyses and describe the data used (e.g., data collection activities, settings, analytic techniques, means of ensuring data quality, comparability of study groups). It is essential for both transparency and the reproducibility of research that researchers adhere to standards that require the reporting of these details.

When data are combined from multiple sources, researchers should verify and report what data elements come from which source, how they are linked, and how these linkages are tested and verified to ensure that data errors do not undermine results (**IR-3**). When data are derived from tests or scales, the test or scale characteristics as well as evaluations of their performance (psychometric properties) should be 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 choices during design and assumptions during data analyses, and these should be declared.

Researchers should describe how they systematically addressed all relevant threats to internal and external validity (Shadish, Cook, and Campbell 2002). Researchers should follow the relevant reporting guidelines established by medical journals and other professional groups. Consistency in reporting makes it easier to evaluate, compare, and synthesize research results (**IR-5**).

Treatment effect estimates can also be biased owing to a lack of masking (also known as blinding). Masking refers to the concealment of the treatment or intervention allocation from one or more individuals involved in a clinical research study. Both randomized controlled trials and observational studies can employ masking as part of the study design. Depending on the nature of the treatment, the type of follow-up required, and/or study resources, it may not always be possible to mask study participants, providers, or investigators. In these cases, researchers should mask the staff involved with collecting and analyzing the data when possible. Lack of masking should be documented in study reports and the potential impact on results discussed (**IR-6**).

Researchers also need to ensure that they adhere to best practices for data management throughout the research process, including developing a data management plan. A DMP should describe the data that will be generated by the study and the documentation requirements and processes that will govern the management, storage, preservation, and potential future uses of study data (**IR-7**). DMPs are fundamental to ensuring the scientific integrity of clinical research, and they also have an additional salutary effect on open science: ensuring that good DMPs are in place at the outset of a study will facilitate data sharing at its conclusion. PCORI, along with several other US funding agencies, now requires DMPs as a condition of research funding (Patient-Centered Outcomes Research Institute 2018; Thoergersen 2015).

4: STANDARDS FOR PREVENTING AND HANDLING MISSING DATA

MD-1: Describe methods to prevent and monitor missing data.

Investigators should explicitly state potential reasons that study data may be missing. Missing data can occur from patient dropout, nonresponse, data collection problems, incomplete data sources, and/or administrative issues. As relevant, the protocol should include the anticipated amount of and reasons for missing data, plans to prevent missing data, and plans to follow up with study participants. The study protocol should contain a section that addresses steps taken in study design and conduct to monitor and limit the impact of missing data. This standard applies to all study designs for any type of research question.

MD-2: Use valid statistical methods to deal with missing data that properly account for statistical uncertainty owing to missingness.

Valid statistical methods for handling missing data should be prespecified in study protocols. The analysis should explore the reasons for missing data and assess the plausibility of the assumptions associated with the statistical methods. The potential impact of missing data on the results and limitations of the approaches used to handle the missing data should be discussed.

Estimates of treatment effects or measures of association should be based on statistical inference procedures that account for statistical uncertainty attributable to missing data. Methods used for imputing missing data should produce valid confidence intervals and permit unbiased inferences based on statistical hypothesis tests. Bayesian methods, multiple imputation, and various likelihood-based methods are valid statistical methods for dealing with missing data. Single imputation methods, such as last observation carried forward, baseline observation carried forward, and mean value imputation, are discouraged as the primary approach for handling missing data in the analysis. If single imputation–based methods are used, investigators must provide a compelling scientific rationale for why the method is appropriate. This standard applies to all study designs for any type of research question.

MD-3: 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 participation in all or only some study activities. Investigators should attempt to continue to collect information on key outcomes for participants unless consent is withdrawn. All participants included in the study should be accounted for in study reports, regardless of whether they are included in the analyses. Any planned reasons for excluding participants from analyses should be described and justified. In addition, missing data owing to other mechanisms (such as nonresponse and data entry/collection) should be documented and addressed in the analyses.

MD-4: Examine sensitivity of inferences to missing data methods and assumptions, and incorporate it into the 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, 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, when possible, the abstract of any reports.

Rationale for These Standards

These standards apply to both missing data and inaccurate data (e.g., in electronic health records), the treatment of which are governed by similar design and analytical considerations (Benchimol et al. 2015). Missing data are unrecorded or inaccurate values or unavailable information that would be meaningful for data analysis and could affect results and conclusions. Possible reasons for missing data include the following:

- Recoding or measurement errors
- Utilizing data sets derived from records not intended for research, such as those generated from routine clinical care
- Involving or evaluating participant populations that are harder to retain over time, making it difficult to collect data

PATIENT VOICES

Sarah

Originally published in 2013

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 regiment, which involved injecting the trial drug, or a placebo, into her abdomen twice daily.

Besides being interested in the benefits she might personally 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, etc., it's very important that you volunteer."

As Sarah began the trial, she found the injections were very difficult to handle. She found the injections to be a painful 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 that 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.

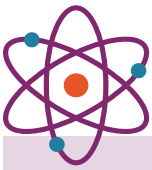
For more about Sarah, see www.healthtalkonline.org/medical_research/clinical_trials/ Topic/3638/Interview/2017/Clip/14719.

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

To address missing or inaccurate data, researchers must have a comprehensive understanding of how the data were generated or collected. These processes should be described to (1) ensure alignment of the approach used to address missing data, the data that are missing, and the causes of missing data and (2) ensure that these processes are clear, reasonable, and can be evaluated by the users of the research. Whether the data are sufficient or the missingness and inaccuracy too great depends on the specific research question(s). There may be cases, particularly with secondary data sources, in which other data sources should be identified for research purposes, given the extent of missingness and/or inaccuracies.

Missing data can occur at two levels: (1) the respondent level ("unit nonresponse"), where an individual chooses not to participate in a study or provide data; and (2) the variable level ("item nonresponse"), where an individual does not answer a specific question or data for a specific variable or time point is not collected. Both types of nonresponse are problematic, although unit nonresponse generally has more impact on the final analyses. Data may not be recorded because of participant actions, such as missing a scheduled follow-up appointment or dropping out of the study altogether. Regardless of the reason the data are missing, if proper statistical methods for handling missing data are not employed, the analyses of those data can be biased or overstate the precision of the findings. These standards do not cover cases called "missing by design," in which data are not available because the study design did not include plans to collect or obtain them.

The issue of missing data is a particularly important consideration in PCOR, given the emphasis on including diverse participant populations and clinical settings. This variety may make collecting complete data sets more challenging. For



RESEARCH IN PRACTICE

Missing Data

Originally published in 2013

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.

example, participants with more than one disease condition or those seen in community care settings may be harder to retain over the course of the study owing to challenges with engagement, trust, access, or other reasons. Preventing missing data is one of several reasons researchers may choose to conduct studies in specialized clinical settings and to exclude participants who may be less likely to complete the study.

Many researchers and groups have provided guidance on how to handle missing data (Li et al. 2014; Little et al. 2012; National Research Council 2010). Rigorous research requires that investigators first identify potential reasons for missing data and include plans to prevent and monitor for missing data in the study protocol (**MD-1**). For example, participants can face various challenges during research studies (see **Patient Voices: Sarah**). Involving patients and other stakeholders (e.g., clinic staff responsible for recruitment and follow-up) during the design of a study can help to identify and address potential reasons for dropout or difficulties in collecting data. Researchers and participants should work together to identify and address those reasons (see **Research in Practice: Missing Data**). The study protocol should justify the choice of statistical methods to handle missing data and describe the underlying assumptions and potential limitations of the methods (**MD-2**). Statistical inference procedures that account for statistical uncertainty owing to missingness—such as Bayesian methods, multiple imputation, and likelihood-based methods—are preferred. Single imputation methods, which fail to account for uncertainty, are discouraged (see **Research in Practice: Bias in Last Observation Carried Forward Method**). The method(s) for addressing missingness should also be selected prior to reviewing the data, to reduce the risk of adversely affecting the validity of the study findings.

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

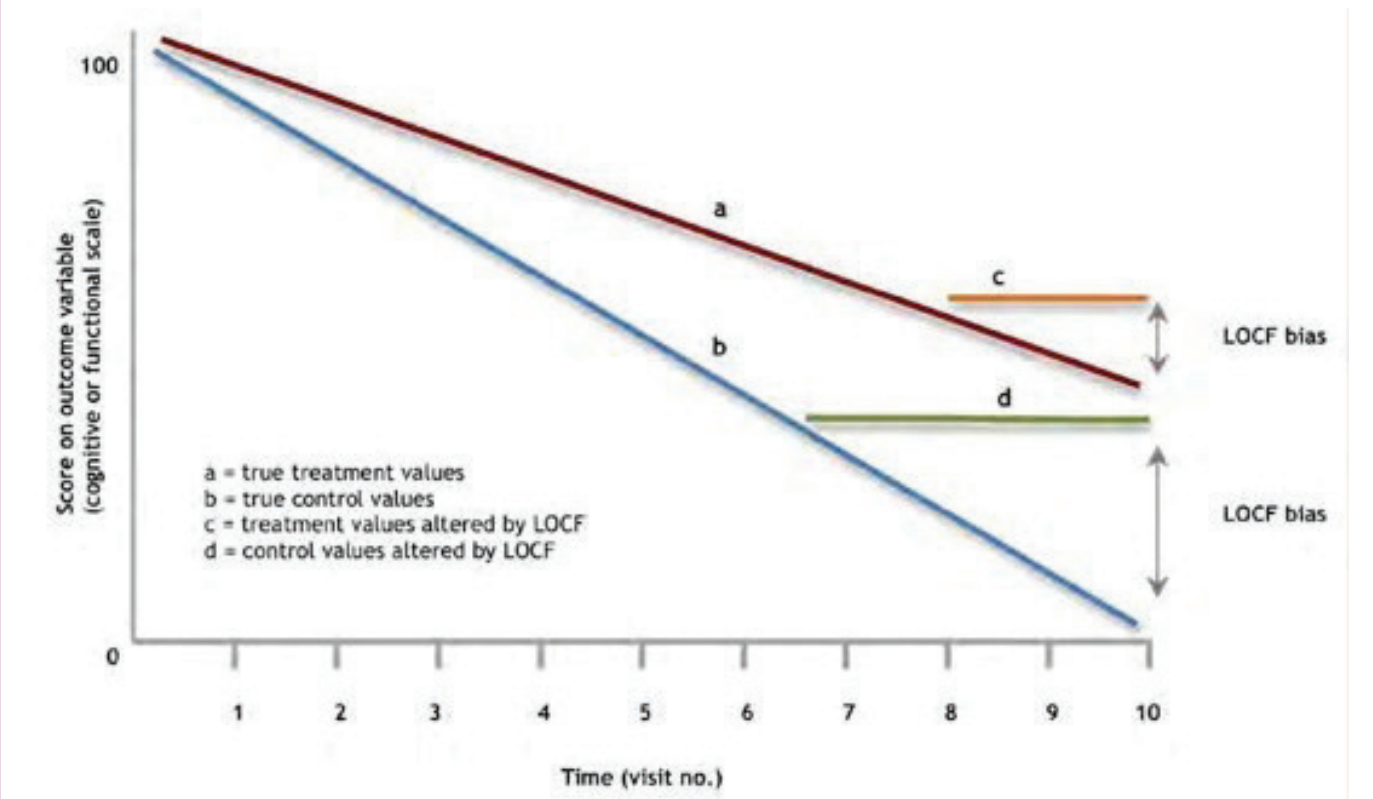
- What is missing has nothing to do with participant characteristics (known as "missing completely at random").
- What is missing depends on participant characteristics predictive of the outcome, and these characteristics were measured ("missing at random").
- What is missing depends on participant characteristics predictive of the outcome, and these characteristics were either not measured or not observed ("missing not at random," or "non-ignorable" missingness).

Investigators should track all study participants, recording when participants drop out as well as the reasons for dropout and attrition (**MD-3**).

Both missing data and the use of inappropriate methods to address missingness can lead to biased findings. Thus, investigators should report the extent and pattern of missing data and conduct a sensitivity analysis (**MD-4**). This analysis will help to determine how the missing data mechanism(s) affect(s) the study results (referred to as assessing the sensitivity of inferences).

RESEARCH IN PRACTICE

Bias in Last Observation Carried Forward (LOCF) Method



For some conditions, such as dementia, patients' cognitive functioning typically worsens over time. That means that a patient assessment collected midway through a trial will overestimate cognitive functioning at the end of it. 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 six 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 above 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 six-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 six months and the true value at 10 months is shown in the figure as the LOCF bias (Molnar et al. 2009).

5: STANDARDS FOR HETEROGENEITY OF TREATMENT EFFECTS (HTE)

HT-1: State the goals of HTE analyses, including hypotheses and the supporting evidence base.

State the inferential goal of each HTE analysis and explain how it relates to the research topic. Specify whether the HTE analysis is hypothesis driven (sometimes denoted as confirmatory) or hypothesis generating (sometimes denoted as exploratory). Hypothesis-driven HTE analyses should be prespecified based on prior evidence (described clearly in the study protocol and study reports), and supported by a clear statement of the hypotheses the study will evaluate, including how subgroups will be defined (e.g., by multivariate score, by stratification), outcome measures, and the direction of the expected treatment effects.

HT-2: For all HTE analyses, provide an analysis plan, including the use of appropriate statistical methods.

The study protocol should unambiguously prespecify planned HTE analyses. Appropriate methods 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. Appropriate methods should be used to account for the consequences of multiple comparisons; 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).

HT-3: Report all prespecified HTE analyses and, at minimum, the number of post hoc HTE analyses, including all subgroups and outcomes analyzed.

Both protocols and study reports must report the exact procedures used to assess HTE, including data mining or any automatic regression approaches. HTE analyses should clearly report the procedures by which subgroups were defined and the effective number of subgroups and outcomes examined. Within each subgroup level, studies should present the treatment effect estimates and measures of variability. Prespecified HTE analyses (hypothesis driven) should be clearly distinguished from post hoc HTE analyses (hypothesis generating). Statistical power should be calculated and reported for prespecified (hypothesis-driven) analyses.

Rationale for These Standards

Because of differences in individual risk factors (e.g., sex, age, co-morbidities, race, lifestyle) and differences in disease stages, people often do not respond the same way to the same treatment. For some, the treatment will produce the intended benefit; for others, the benefit may be less than what was intended. Yet in others, the treatment may have no effect or have harms that outweigh the benefits. Heterogeneity of treatment effect is the technical term used to describe this variability in treatment responses.

Patient-level information about the benefits and harms of a treatment is not always well described in research reports. Variations in responses to a treatment can be masked by study design and analysis. Clinical trials and observational studies often report only the average treatment effects (i.e., the effect of a treatment averaged across all study participants). Failure to measure and/or appropriately analyze variables that could be used to predict different treatment responses can also make it difficult to determine the effect of a treatment for a specific type of patient.

Explicitly addressing HTE in clinical research helps to 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**). The importance of understanding individual variability and how it affects the prevention and treatment of disease is a core tenet of "personalized" or "precision" medicine initiatives (Dahabreh, Hayward, and Kent 2016).

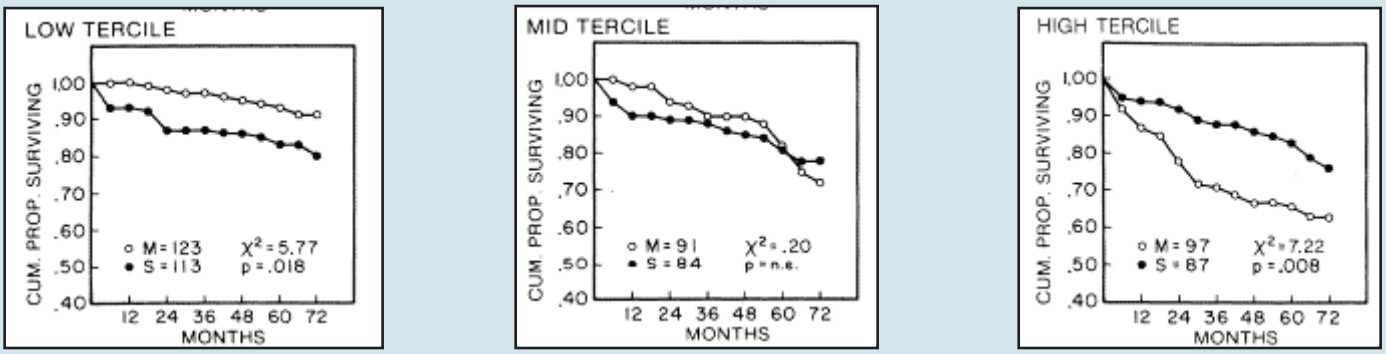
Methods to assess HTE vary in terms of methodological sophistication as well as the extent to which they can generate valid and reliable estimates of treatment effects. The central challenge of HTE analyses is to improve the patient-level information about the risks and benefits of a treatment while minimizing the possibility of spurious conclusions—namely, falsely detecting HTE (referred to in statistics as Type I error) or failing to detect true HTE (Type II error) in particular patient groups (PCORI 2016).

HTE analyses 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. (This first type of approach to HTE is covered by these standards.) The most common approach is to use subgroup analyses to estimate the effects of treatments in a

Heterogeneity of Treatment Effects

The figures below show six-year survival rates during the 1970s for patients with chest pain (angina) at risk for mortality from heart disease. Patients were randomly assigned to heart bypass surgery (black dots) or a nonsurgical 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 et al. (1981), reprinted by permission of Wolters Kluwer Health, provided by Copyright Clearance Center.

specified subset of the study participants. Prediction of individual effects is less common, although it is of increasing interest given the growth in the field of personalized medicine and advances in decision analytic and simulation methods for developing clinical prediction models (Kent, Steyerberg, and van Klaveren 2018).

To estimate the effect of treatment separately for patient groups, researchers often stratify by subgroup (i.e., performing the analysis for just one group of participants, such as women). However, this approach is susceptible to the well-known problem of multiple post hoc analyses that can yield an increased likelihood of Type I or Type II errors. Although estimating stratified treatment effects may be valid for testing a limited number of subgroups when sample sizes are large enough, this approach is inappropriate for inferring HTE when multiple subgroup comparisons are required. An alternative to "one-at-a-time" variable analysis is to conduct a risk-stratified analysis using multivariate prediction tools, which can simultaneously account for multiple risk factors and improve the statistical power of the analysis (Kent et al. 2010).

The first step to ensuring high-quality HTE analyses is to understand the purpose of the research; therefore, the standards require that researchers state the goals for HTE analyses (**HT-1**). Researchers should consider the sample size, data quality, and available evidence and determine whether the analysis is hypothesis driven (sometimes denoted as confirmatory) or hypothesis generating (sometimes denoted as exploratory). The designation (and justification) for all HTE analyses should be made clear to ensure the appropriate design and analysis plan for the study and to allow stakeholders

to interpret results correctly.

HTE analyses should be conducted in accordance with well-defined analytical plans and employ the use of appropriate methods (**HT-2**). First, 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 (Brookes et al. 2001; Goldfine, Kaul, and Hiatt 2011; Lagakos 2006). Second, assessing HTE requires the use of appropriate statistical contrasts (e.g., interaction tests, estimates of differences in treatment effects estimates with standard errors, Bayesian shrinkage estimates). A common error in HTE analyses is to claim differences in treatment effect when one subgroup shows a statistically significant treatment effect and another does not. 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. These requirements apply to both randomized trials and observational studies. Although patients are randomized to the treatment arms in RCTs, subgroups are not randomized, resulting in subgroups with different baseline characteristics, which may confound the interpretation of results.

Protocols and study reports should provide sufficient detail regarding all HTE analyses that were conducted, including the procedures used to assess HTE, selection of outcomes, and effect estimates (**HT-3**). Failure to adequately report on HTE analyses undermines the transparency of the research process and makes it difficult to ensure that findings are appropriately interpreted and applied in practice.

6: STANDARDS FOR DATA REGISTRIES

DR-1: Requirements for the design of registries

Registries established to conduct patient-centered outcomes research must have the following characteristics:

A. Registry Purpose and Protocol. The purpose of the registry should be clearly defined to guide the design of key registry features, including, but not limited to, the target population, the research question(s) to be addressed, the data source used, the data elements collected, data-sharing policies, and the stakeholders involved in developing and using the registry. Participants and other key stakeholders should be engaged in registry design and protocol development. Registries should aim to be user oriented in design and function.

B. Data Safety and Security. Registry custodians should comply with institutional review board (IRB) human subjects protection requirements, the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, and all other applicable local, state, and national laws. Registries should provide information that describes the type of data collection (primary or secondary source data), data use agreements, informed consent documents, data security protections, plans to maintain data protection if the registry ends, and approaches to protect privacy, including risk of and/or process for re-identification of participants, especially for medical or claims records.

C. Data Elements and Quality. Standardized data element definitions and/or data dictionaries should be used whenever possible. When creating a new registry, published literature should be reviewed to identify existing, widely used definitions of outcomes, exposures, and confounders before new definitions are drafted.

When collecting primary data, conduct multistakeholder engagement with potential participants and data users to prioritize data collection needs. When participants support their face validity, use validated instruments or PRO measures when available. If secondary data sources (e.g., electronic medical records, claims data) are used, describe the original purpose of the secondary data and verify the accuracy and completeness of the data, as well as the approach to and validity of the linkages performed between the primary and secondary sources.

The specifics of the quality assurance plan will depend on the type of data (primary or secondary) collected by the registry. In general, the plan should address (1) structured training tools for data abstractors/curators; (2) the 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 (where feasible and appropriate), and validation statistics focused on data quality for the key exposure and outcome variables and key covariates. A risk-based approach to quality assurance, focused on variables of greatest importance, is advisable.

D. Confounding. During the planning phase, registries should identify important potential confounders pertinent to the purpose and scope of the research. During the analysis phase, they should collect reasonably sufficient data on these potential confounders to facilitate the use of appropriate statistical techniques. When conducting analyses, refer to the PCORI Methodology Standards for Data Integrity and Rigorous Analyses and Standards for Causal Inference Methods.

E. Systematic Participant Recruitment and Enrollment. Develop a sampling plan of the target population and identify recruitment strategies for participants that minimize the impact of selection bias. Participants should be enrolled systematically, with similar procedures implemented at all participating sites and for each intervention of interest. Confirm adherence to agreed-on enrollment practices.

F. Participant Follow-Up. The objective(s) of the registry should determine the type, extent, and length of participant follow-up.

Describe the frequency with which follow-up measures will be ascertained, consider linkage with other data sources (e.g., the National Death Index) to enhance long-term follow-up, and identify the date of last contact with the participant in existing registries, where appropriate. Ensure that the participants are followed in as unbiased a manner as possible, using similar procedures at all participating sites.

Monitor loss to follow-up to ensure best efforts are used to achieve follow-up time that is adequate to address the main objective. At the outset of the registry, develop a retention plan that documents when a participant will be considered lost to follow-up and which actions will be taken to minimize the loss of pertinent data. Retention efforts should be developed with stakeholders to ensure that the efforts are suitable for the target population

and that anticipated challenges are addressed appropriately.

DR-2: Documentation and reporting requirements of registry materials, characteristics, and bias

Clearly describe, document with full citations where appropriate, and make publicly available registry materials, including, but not limited to, registry protocols, data-sharing policies, operational definitions of data elements, survey instruments used, and PROs captured. Modifications to any documents or data collection instruments should be clearly described and made available for registry users and participants. Characteristics of the registry participants should be described. To help assess potential selection biases, identify how the participants may differ from the target population. Document the loss to follow-up and describe the impact on the results, using sensitivity analyses (prespecified where possible) to quantify possible biases. Report the extent of bias clearly to stakeholders who may want to use the registry resource.

DR-3: Adapting established registries for PCOR

Previously established registries that intend to support new clinical research may not have been informed by all applicable methodology standards. When new research will use such registries, investigators should engage key stakeholders, including registry participants, to assess the feasibility of using the registry for new research and to ensure the following:

- Informed consent documents are appropriately tailored to participant needs, characteristics, and conditions.
- Data elements are meaningful and useful to researchers and participants.
- Recruitment and retention strategies are feasible and effective.
- Registry policies are patient centered and the use of registry data is transparent to participants.
- Dissemination practices are appropriate and effective at reaching the communities from which the data are collected.
- Opportunities for bidirectional benefit exist between participants and researchers.
- Registry materials, described in **DR-2**, and informed consent forms are publicly available in accessible formats.

DR-4: Documentation requirements when using registry data

Researchers planning PCOR studies that rely on registries must ensure that these registries meet the requirements contained in **DR-1** and **DR-2** and must document each required feature of each registry to be used (e.g., in an appendix to the funding application or study protocol). Deviations from the requirements in **DR-1** and **DR-2** should be well documented and limitations of research related to the deviations from requirements should be addressed when reporting study findings.

Rationale for These Standards

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 treatment or with the defining characteristic of the patients (e.g., device implantation, familial cancer risk).

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. Registries can also be used to prompt or require the collection of additional data about a group of patients with a specific condition (e.g., diabetes, cancer), who undergo a diagnostic test (e.g., a positron emission tomography [PET] scan), or who 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 different treatments. For example, collecting postoperative data about a group of patients who had hip replacements allowed researchers to uncover a significant problem with one type of artificial hip (see **Research Stories: National Joint Registry of England and Wales**).

When registries are properly designed (Agency for Healthcare Research and Quality 2016), 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. The fact that registries are based on medical care as it is delivered in real-world situations increases the likelihood that the findings will be broadly applicable (see **Research in Practice: Data Registries**).

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 show that metal-on-metal hip replacements are more likely to fail than metal-on-ceramic or ceramic-on-ceramic products in the five years after hip surgery.

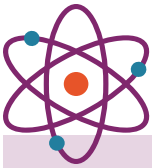
A 60-year-old man undergoing a primary hip replacement with a relatively small (28-millimeter-diameter) ceramic-on-ceramic product can expect a 2.0 percent risk of product failure during the first five years, while the same man with a similar metal-on-metal product can expect a 3.2 percent risk of product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

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

Although registries reflect real-world clinical practices, such data also have limitations for informing healthcare decisions. Data derived from clinical sources often may not meet the same level of quality control as data collected in a clinical trial or even some prospective cohort studies (Brennan and Stead 2000; Kahn, Eliason, and Bathurst, 2010). The methods of collection, definitions of data elements, and interpretation of data about treatments, diseases, and care pathways 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. Researchers who design studies based on registries need to understand the data and ensure the data's quality and relevance to their study. Furthermore, registry data analysis needs to formally consider other influences on outcomes (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 and the data need to be used appropriately.

The quality of data derived from registries depends on a wide array of factors, including design, data elements, data sources, governance, and maintenance. Similar to other research that uses patient health data, registries must be carefully planned, and oversight is needed to prevent confidentiality breaches. Because registries typically follow a patient's natural history, they require multiple follow-up points. Registries are often most useful when they are maintained with data collected in a consistent way over periods that are long enough to capture long-term outcomes that are important to patients (see **Patient Voices: Suzanne)**. However, the problem of missing data may be significant in registry studies requiring long-term data collection that includes multiple patient contacts.

DR-1 specifically addresses the design and maintenance of registries. Registries are most likely to generate valid and relevant findings 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 consider the possibility of re-identification. Planning how best to collect and aggregate the data, ensure data security and the protection of patient privacy, ensure data quality and systematic participant recruitment and enrollment, and track follow-up increases the likelihood that the registry can answer essential PCOR questions. Once the registry is established, researchers should clearly document and report on the registry's materials, characteristics, and potential sources of bias to ensure transparency to stakeholders who may want to use the registry data and/or results (**DR-2**). Researchers are encouraged to make registry information publicly available by submitting registry profiles to centralized, publicly accessible depositories, such as the Registry of Patient Registries maintained by



RESEARCH IN PRACTICE

Data Registries

Originally published in 2013

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 et al. 2011).

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 from 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.

the Agency for Healthcare Research and Quality (AHRQ).

Expanding the scope of an established registry to answer PCOR questions provides an opportunity to leverage existing resources to address a broader set of stakeholder needs (**DR-3**). When undertaking such efforts, stakeholder engagement can ensure that appropriate patient-centered adaptations are considered, including reevaluating key informed consent documents when new research questions arise, expanding the collection of data elements and outcomes to include those most meaningful to participants, launching additional recruitment strategies that are realistic and feasible for participants, and optimizing dissemination practices to ensure that results reach all relevant communities participating in the registry efforts.

Researchers need to consider the same elements of the registry that were considered when it was designed; however, they also need to consider the advantages and limitations of the registry's data for their particular research question. Researchers must 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). Finally, researchers planning PCOR studies that rely on registries must meet documentation requirements for the registry being used and report any deviations from the previous standards along with study findings (**DR-4**).

PATIENT VOICES

Suzanne

Originally published in 2013

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 (i.e., 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.

7: STANDARDS FOR DATA NETWORKS AS RESEARCH-FACILITATING STRUCTURES

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

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

A. **Data Integration Strategy.** 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 (including code and process documentation) 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-identifying data and apply algorithms to ensure that the desired level of confidentiality is achieved to meet the particular PCOR application's need. Data custodians should ensure that the data privacy/consents of the original data source cover the intended data usage. Privacy protections, including which data will be released and how breaches will be addressed, should be specified in the data use agreement. The physical security of the data and data platforms should be considered and addressed as well.

C. **Identity Management and Authentication of Individual Researchers.** Develop reliable processes to verify and authenticate the credentials of researchers who are granted access to a distributed research network.

D. **IP Policies.** A research network should develop policies for handling and disseminating 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 that uses the network. Guidelines should balance (1) minimizing impediments to innovation in research processes and (2) making the research results widely accessible, particularly to the people who need them the most.

E. **Standardized Terminology Encoding of Data Content.** The data content should be represented with a clearly specified standardized terminology system to ensure that its meaning is unambiguously and consistently understood by parties using the data.

F. **Metadata Annotation of Data Content.** Semantic and administrative aspects of data content 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 organized into a standard structure that establishes common definitions and shows close or distant associations among variables. A common data model specifies necessary data items that need to be collected and shared across participating institutes, clearly represents the 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 that rely on data networks must ensure that these networks meet the requirements contained in **DN-1**, and they must document the current maintenance status of the data network (e.g., currency of the data, level of data curation). Because different studies are expected to have different dependencies on various components of the data network, researchers should assess the appropriateness of the data in the network for a specific research study through the following activities:

A. **Data content and conformance.** Document what is actually needed for the research question and compare that to the sources in the network. Identify which data are best represented by the network's data sources and how they are included in the study. Ensure that the representations and values of the data to be used from the network are sufficient to address the research question.

B. **Data quality.** Assess the data quality for the data sources that will be used. It is especially important to assess data completeness and plausibility. Where data are incomplete, identify and assess potential biases for completeness and consider alternate sources. Assess plausibility by reviewing data value distributions and comparing additional data sources that would have expected concordance with the selected sources.

Determine whether the data sources are of sufficient quality to be included in the analysis.

C. **Sensitivity analyses.** After the initial analysis is completed, perform sensitivity analyses on the data sources to test whether possible variations in data characteristics would affect the conclusions of the analysis. Specifically, measure the sensitivity of the conclusions to the following:

- Completeness and correctness of the data in the data network
- Availability of data sources that are most likely at risk of exclusion
- Temporal dependence of the data
- Operational definitions and decisions made to implement analysis

The results of these assessments should be documented and included with any findings from research studies that use the data networks.

Rationale for These Standards

Collaborative data networks are agreements that coordinate data use across healthcare organizations. Data networks aggregate information from a range of data types (e.g., claims, medical records, pharmacy records, lab/pathology reports) and/or from various medical settings (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 and observational studies. Data networks designed to facilitate research include such key components as 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. Data networks may cover a wide range of potential research topics, such as studying the effectiveness of diagnostic tests, monitoring adverse effects of new drugs or devices, and testing new cancer treatments.

Data networks have many characteristics that make them important for the development and advancement of PCOR. Analyzing data already collected across organizations or medical settings can be 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 from any one site. This variety means that the results are more likely to be generalizable, improving the relevance of information to patients and clinicians.

Data networks are also more likely to 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 (Goodman, Schneeweiss, and Baiocchi 2017), it can make it possible to detect smaller differences in outcomes or recognize differences in less time. Large numbers of records make it easier to determine whether the comparative effectiveness of a treatment varies across subgroups (e.g., between men and women, among people with different co-morbidities).

Despite 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. Setting standards for data networks ensures that key components are included when networks are designed—and that these components are considered when data from these networks are used in research studies.

Several organizations in the United States, Canada, and Europe have developed guidelines, identified best practices, and supported 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 HIPAA—or general policies with applications in health care, such as the Organisation for Economic Cooperation and Development personal privacy guidelines (OECD 2013). A detailed discussion of all existing guidance is beyond the scope of this report, but investigators who conduct research on data networks should be familiar and comply with applicable laws, institutional policies, and additional methodological guidance.

The PCORI Methodology Standards for Data Networks recognize that the construction and management of the network is separate from the use of network data for PCOR studies. The first standard addresses the development and maintenance of a network's policies and procedures, and it specifies key elements necessary for a successful network that will generate useful data (**DN-1**). Definitions and other characteristics of data elements need to be clear, agreed on, and verified.

Processes need to be created and documented for the transformation of data elements so they are equivalent even when they come from different sources. Creating and maintaining standardized terminology (Kahn et al. 2016) and data descriptions require planning and resources.

Data networks link and share information about individuals in ways that could compromise patient privacy. 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 are also necessary. Generally, study proposals and protocols should describe data use agreements, informed consent, and approaches to data security. Proposals should also describe how researchers will address the risk of re-identifying patients and how the actual use of data compares 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 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.

A data network's usefulness often increases with its longevity. 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 that result from the network (i.e., IP policies). Standardized terminology is necessary, and information about the data elements (known as metadata) must be provided. Data elements should also be assembled into a model that shows the relationships among the data elements and helps all users to interpret the data correctly (Kahn, Batson, and Schilling 2012).

The second standard (**DN-2**) addresses the activities of researchers who seek to access and use data from an existing network. Increased availability of large volumes of data ("big data") have raised concerns that data availability, rather than data suitability, are driving the use and analysis of this information in PCOR studies. Because the appropriateness of a data source varies according to the specific research question and how the data are used, the content and quality of data in a network cannot be certified for appropriateness for all research questions. Therefore, assessments must be conducted as part of individual research studies.

Important categories of data content and quality have been identified as conformance, completeness, and plausibility (Kahn et al. 2015). To identify potential threats to data validity, these categories should be assessed specifically for research data derived from secondary sources, including verifying that data values returned by queries reflect what was expected. Data equivalence evaluation for all involved data sources against each other should be documented, and any limitations should be clearly outlined.

Because the assessments of content and quality are often qualitative, sensitivity analyses should be used to provide some measurement of how the specific data vulnerabilities may become threats to the research's validity. Quality assurance measures of the data sources should be assessed and documented. Any limitations imposed on the data network owing to quality limitations of single data sources should be evaluated and documented.

8: STANDARDS FOR CAUSAL INFERENCE METHODS

CI-I: Specify the causal model underlying the research question *(cross-cutting standard, applies to all PCOR/CER studies)*. Researchers should describe the causal model relevant to the research question, which should be informed by the PICOTS framework: populations, interventions, comparators, outcomes, timing, and settings. The causal model represents the key variables; the known or hypothesized relationships among them, including the potential mechanisms of effect; and the conditions under which the hypotheses are to be tested. Researchers should use the causal model to determine whether and how the study can handle bias and confounding and the extent to which valid estimates of the effects of an intervention can be generated given the particular hypothesis, study design, analytical methods, and data source(s).

CI-2: Define and appropriately characterize the analysis population used to generate effect estimates. Researchers should specify the eligibility criteria for inclusion in the study population and analysis. In prospective studies, decisions about which patients are included in an analysis should be based on information available at each patient's time of study entry; in retrospective studies, these decisions should be based on information from a defined period before the start of exposure. For time-varying treatment or exposure regimes, specific time points should be clearly specified; relevant variables measured at baseline and up to, but not beyond, those time points should be used as population descriptors. 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), address selection bias that may be introduced by excluding patients, and assess the potential impact on the validity of the results.

CI-3: Define with the appropriate precision the timing of the outcome assessment relative to the initiation and duration of exposure. To reduce potential sources of bias that arise from inappropriate study design choices (e.g., immortal time bias), researchers must precisely define, to the extent possible, the timing of the outcome assessment relative to the initiation and duration of the exposure.

CI-4: Measure potential confounders before the start of exposure and report data on potential confounders with study results. In general, variables used in confounding adjustment (either in the design or analysis) should be ascertained and measured before the first exposure to the intervention(s) 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 adjust for measured confounding, researchers should consider and report how propensity scores will be created (high dimensional propensity score versus a priori clinical variables) and which balancing method will be used (e.g., matching, weighting, stratifying). Researchers should assess and report 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. When an instrumental variable (IV) approach is used (most often to address unmeasured confounding), empirical evidence should be presented that describes how the variable chosen as an IV satisfies the three key properties of a valid instrument: (1) the IV influences the choice of 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).

Rationale for These Standards

One of the key objectives of health research 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

RESEARCH STORIES

Human Immunodeficiency Virus

What is the optimal time for patients with HIV infection to start combined antiretroviral therapy? 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 effectiveat 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 (the HIV-CAUSAL Collaboration 2011).

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 are a methodological answer to this problem. Because RCTs 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. If a similar distribution of potential confounders across all the different possible assignments of patients were achieved, then the average estimate of how much the intervention affects the outcome would be correct, even if individual participants differ in ways other than the treatments they receive.

The problem is that not all questions can be studied using a randomized trial, and even when they can, randomization cannot address all threats to the validity of results. Researchers often 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—for settings in which a randomized trial is impossible, unethical, or too costly. But even in randomized trials, post-randomization confounding or selection bias may occur (from, for example, informative patient dropout, crossover to other treatments, protocol violations), or randomization may produce groups that are different in important ways by chance.

By helping to address sources of confounding and bias from design-related errors, causal inference methods focus on increasing confidence that the treatment being studied is causing the outcome (see **Research Stories: Human Immunodeficiency Virus**). Methods to address confounding include various forms of population restriction and regression methods. Each method also addresses the issue of confounding differently. For example, propensity scores, like standard regression methods, cannot directly solve the problem of unmeasured confounding factors, but they can adjust for multiple confounders and variables that serve as proxies for other, unmeasured confounders (Rosenbaum and Rubin 1984). IV methods, on the other hand, purport to get around the unmeasured confounder problem by identifying and exploiting naturally occurring distributions of treatment choices that resemble randomization but that rely on additional assumptions that are untestable using the data available. While these tools are both powerful and useful, they have important limitations. Most of these methods can control only for the effect of confounders that are actually identified (and for which data are available). The assumptions made in any of these methods also require extraordinarily close scrutiny.

Although these statistical methods can produce more accurate estimates of treatment effects and uncertainty, none address serious threats to valid causal inference arising from design-related errors, including selection bias, reverse causation, and adjustment for intermediate variables (Goodman, Schneeweiss, and Baiocchi 2017). More broadly,

sophisticated analytical methods cannot compensate for poor design or low-quality data. Therefore, the Standards for Causal Inference Methods should be understood as applying to both the design and the analysis of observational studies, with the exception of **CI-1**, which applies to all PCOR studies, including RCTs.

Researchers should always begin by explicitly articulating the hypothesized causal model underlying the research question and detailing how the study is designed to assess the particular effect(s) of interest (**CI-1)**. The appropriate application of analytical methods and interpretation of results depends on the specification of a causal model, study design, and causal relationship(s) of interest (Petersen and van der Laan 2014).

Observational studies should be designed to emulate an RCT (Goodman, Schneeweiss, and Baiocchi 2017; Hernán et al. 2008), which requires specifying the eligibility criteria for inclusion in the study population and analysis (**CI-2**) and clearly defining the timing of the outcome measurement relative to the treatment or exposure (**CI-3**). 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. An alternative is to employ a new-user design, which restricts the analysis to new (rather than prevalent) users of a treatment and the appropriate comparison group (Ray 2003).

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**). Whether a variable is treated as a confounder should be based on subject matter knowledge and the underlying causal model. Adjusting for variables that are not confounders, including intermediate variables (mediators), can introduce additional bias (Schisterman, Cole, and Platt 2009).

Creating standards specific to all current statistical methods for causal inference that are applicable to all potential research questions is not feasible; the choice of appropriate statistical methods depends on the research question of interest, including the causal relationship of interest, and the data source(s) utilized. Given this situation, standards are included for two general types of analysis that are relatively well developed and increasingly used in PCOR: propensity scores (**CI-5**), which can be used to address measured confounding, and instrumental variables (**CI-6**), which can be used to address both measured and unmeasured confounding, but with untestable assumptions. When any sophisticated analytical approaches are used, transparency is particularly important. Sensitivity analyses are also critical, and additional efforts are required to document the assumptions underlying the analyses and how these assumptions were examined.

9: STANDARDS FOR ADAPTIVE AND BAYESIAN TRIAL DESIGNS

AT-1: Specify planned adaptations, decisional thresholds, and statistical properties of those adaptations.

The adaptive clinical trial design must be prospectively planned and the design must be clearly documented in the study protocol before trial enrollment begins, including at a minimum the following:

- All potential adaptations, including timing
- Interim trial findings that will be used in determining each adaptation
- Statistical models and decisional thresholds to be used
- Planned analyses of the trial endpoint(s)

The description of the design should be sufficiently detailed that it could be implemented based on the description of procedures. This specification should include a statistical analysis plan in which all necessary detail is provided regarding planned interim and final analyses.

Additionally, the statistical properties of 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, 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.

AT-2: Specify the 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 about how the prior distribution was determined and if an informative or noninformative 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. Computational issues should be addressed, including describing the choice of software and creating, testing, and validating that software. Software used for Bayesian calculations during the trial's design, execution, and final analysis must be functionally equivalent. When feasible, software or other computing packages should be made available to relevant stakeholders for evaluation and validation.

AT-3: Ensure that clinical trial infrastructure is adequate to support planned adaptation(s) and independent interim analyses.

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 because this information provides data regarding treatment success.

AT-4: When reporting adaptive randomized clinical trials, use the CONSORT statement, with modifications.

The following sections of the 2010 CONSORT statement can be used to report key dimensions of adaptation:

- Adapting randomization probabilities (sections 8b and 13a)
- Dropping or adding study arms (sections 7b and 13a)
- Interim stopping for futility and superiority or adverse outcomes (sections 7b and 14b)
- Reestimating sample size (sections 7a and 7b)

- Transitioning stages (e.g., seamless Phase II/III designs; sections 3a, 7a, 7b, and 16)
- Modifying inclusion and exclusion criteria (sections 4a and 13a)

CONSORT sections 16, 20, and 21 provide additional guidance on reporting aspects of an adaptive trial.

All possible adaptations included in the prospective design, even if they did not occur, should be included in the study reports.

Rationale for These Standards

Randomized trials have advantages and disadvantages in determining the comparative effectiveness of different interventions. RCTs can provide strong evidence, but they are also often perceived as taking too long to produce results or being too rigid in a rapidly changing field. One solution is to employ adaptive trials, which build on the approaches used in most clinical trials but differ in that they allow changes to be made to a study after it has begun. An adaptive clinical trial is one in which key trial characteristics (e.g., randomization proportions, sample size, treatment arms, eligibility criteria) evolve according to prespecified rules during the trial in response to information that accrues within the trial itself. Potential advantages of this approach include statistical efficiency, improved patient outcomes, or improved balance of risks and benefits to trial participants (Berry et al. 2010). Rather than waiting until the end of the study period to see the results and suggest changes for the next study, changes are planned as part of the trial design and executed based on the analyses conducted during the trial.

Recognizing the need for innovative clinical trial design, representatives from the National Institutes of Health'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, specialized expertise and experience are required to design and conduct these trials.

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 reestimation—that have become standard practices. Many adaptive features can be implemented individually using classical statistics, often called frequentist approaches, but complex designs that combine several dimensions of adaptation typically require a different statistical approach known as Bayesian analyses. These adaptive designs allow for the incorporation of prior or external information that may be similar to, but not exchangeable with, information in the proposed trial.

Adaptive trials should adhere to the principles of good design and analysis that apply to all rigorous research; however, their complexity can make this more difficult, requiring extra attention to specific steps in the research process. The experience in therapeutics and device trials, combined with theoretical considerations, provide the basis for standards that govern the design and conduct of adaptive trials in PCOR. Additional guidance is available in the published literature, including an FDA draft guidance document on this topic (US Food and Drug Administration 2010a).

Good adaptive trial design requires preplanning and the specification of procedures at the outset. Adaptive trials typically require that simulations or sensitivity analyses be conducted during the design phase to define the error rates. Descriptions of the design— in both protocols and published papers—must include adequate detail about the study elements and planned adaptations. Given the potential complexity introduced by adaptations, the timing of the interim analyses and the changes that could be made based on those data should be determined before the trial starts (**AT-1**). In addition, adaptive trials that use Bayesian approaches require even more detailed specification of the analysis plan than is typically provided 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-2**).

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-3**). Once an adaptive trial is complete, the 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-4**).

10: STANDARDS FOR STUDIES OF MEDICAL TESTS

(Formerly Standards for Studies of Diagnostic Tests)

MT-1: Specify clinical context and key elements of the medical test.

Evaluation of tests used to inform medical decision making (e.g., diagnostic tests, prognostic tests, predictive tests) should specify each of the following items and provide justification for 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 choice of comparator (e.g., another test, no test) and goal of the comparison; (3) the technical specifications of the test(s) as implemented in the study; (4) the approach to test interpretation; (5) the sources and process for obtaining reference standard information, when applicable; (6) the procedures for obtaining follow-up information and determining patient outcomes, when applicable; and (7) the clinical pathways involving the test and the anticipated implications of test use on downstream processes of care and patient outcomes. These items ought to be specified for all types of tests used for medical decision making and 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.

MT-2: Assess the effect of factors known to affect performance and outcomes.

Studies of tests used to inform medical decision making should include an assessment of the effect of important factors known to affect test performance and outcomes, including, but not limited to, the threshold for declaring a "positive" test result, the technical characteristics of the test, test materials (e.g., collecting, preparing, and handling samples), operator dependence (e.g., lab quality, interpretation requirements), and the care setting.

MT-3: Focus studies of medical tests on patient-centered outcomes, using rigorous study designs with a preference for randomized controlled trials.

A prospective randomized design should be used when possible to assess the diagnostic, prognostic, predictive, and/or therapeutic outcomes of testing. If a nonrandomized design is proposed, a rationale for using an observational study (or modeling and simulation) should be provided, and efforts to minimize confounding documented.

Rationale for These Standards

Medical tests—which include a broad range of chemical, imaging, electrical, functional, and visual examinations—are an essential part of modern medicine. Healthcare providers recommend tests to screen for unrecognized conditions, test diagnostic hypotheses, estimate the location or extent of a disorder, develop prognostic estimates, or measure treatment response. Patients, caregivers, and clinicians need specific information about the expected benefits and harms of a test in their particular circumstances when deciding whether a test should be performed. When the research on a test is flawed, clinicians may underestimate or overestimate the likelihood that a patient has (or is at risk of developing) a disease and thereby provide misleading information to patients and caregivers. Medical tests may also expose patients to unnecessary inconvenience or harm, including radiation exposure and complications from invasive procedures undertaken in response to test results.

Overall, the impact of medical testing on patient outcomes has often been understudied in clinical research. Although these tests generate information, they do not necessarily (or directly) produce a better outcome for the patient. Studies of medical tests tend not to assess all relevant effects on patients, particularly long-term benefits and harms, as well as cognitive, emotional, social, and behavioral effects (Bossuyt and McCaffery 2009). To improve patient outcomes, the test results 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 medical 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.

Medical tests can be 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 medical tests (Lord, Irwig, and Bossuyt 2009). Although guidelines exist that address the reporting of diagnostic or predictive accuracy studies, standards have not been well defined for studying the impact of medical tests on subsequent care or patient outcomes (see the Standards for Data Integrity and Rigorous Analyses for more information on reporting guidelines).

The standards for studies of medical tests reflect three principles for rigorous PCOR. The first standard emphasizes the

importance of understanding key elements of medical tests and the clinical context in which the test is used (**MT-1)**.

The second standard asserts that accuracy alone is often not a sufficient measure of the benefit of a test. The overall scientific validity and clinical utility of a medical test depend on knowing how key factors affect clinical outcomes (Ferrante di Ruffano et al. 2012). Studies should include an assessment of the effect of factors known to affect test performance and outcomes, including the threshold for declaring a "positive" test result, the technical characteristics of the test, test materials (e.g., collecting, preparing, and handling samples), operator dependence (e.g., lab quality, interpretation requirements), and the care setting (**MT-2**).

The third standard underscores how alternate tests or testing strategies should be compared in terms of their effects on patient-centered outcomes using the optimal and most feasible study design (**MT-3**). Although a randomized study designed to capture relevant patient outcomes generally provides the strongest clinical evidence, the use of RCTs is not always feasible; alternative approaches to performing clinical studies of medical testing are appropriate in some situations (Lord, Irwig, and Bossuyt 2009). When nonrandomized designs are used, the choice of study design should be justified and strategies for minimizing the risk of bias in the nonrandomized design described. Regardless of study design, investigators should ensure that important patient-relevant outcomes are accounted for in the study.

11: STANDARDS FOR SYSTEMATIC REVIEWS

SR-1: Adhere to National Academy of Medicine (NAM) standards for systematic reviews of comparative effectiveness research, as appropriate.

Systematic reviews, which critique and synthesize the existing literature, can also identify evidence gaps and inform decisions of how to address these gaps. Existing standards for systematic reviews developed by credible authorities, such as the Cochrane Collaboration and the Agency for Healthcare Research and Quality, vary somewhat in their recommended approaches. The PCORI Methodology Committee endorses the standards issued by NAM in 2011 but recognizes both the importance of conducting systematic reviews consistent with updates to best methodological practices and that there can be flexibility in the application of some standards without compromising the validity of the review, including the following:

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

Researchers should describe and justify any departures from the 2011 NAM standards (e.g., why a particular requirement does not apply to the systematic review).

Rationale for These Standards

Systematic reviews find, assess, and synthesize results from several individual studies 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. Systematic reviews are important for PCOR because they facilitate the efficient use of existing research results and aid in targeting future research. Often, it is only by looking at a large body of evidence that it is possible to assess the comparison of different health interventions (see **Research Stories: Getting off the Ventilator**).

Systematic reviews also make it possible to determine which relevant patient-centered 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**).

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, and the criteria used to include or exclude studies as well as how the included studies are evaluated may differ. Results may also be affected by errors when data are collected and combined from different studies.

In 2011, the National Academy of Medicine (then known as the Institute of Medicine) released a report titled *Finding What Works in Health Care: Standards for Systematic Reviews* (Institute of Medicine 2011). PCORI has concluded that these standards are generally useful, although emerging literature and methods may augment these standards for use in PCOR.
The NAM standards were developed by an expert panel based on a broad review that considered and incorporated existing authoritative sources (e.g., Cochrane Collaboration, AHRQ Evidence-Based Practice program). The NAM standards are designed to support consistent application of a well-defined set of methods 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 NAM standards are appropriate for inclusion in the PCORI Methodology Standards because they aim to ensure patient centeredness in conducting systematic reviews of clinical effectiveness research (**SR-1**).

The NAM standards address how to design and conduct systematic reviews that rely on published data and conventional statistical models; however, they do not address network meta-analysis or individual participant data meta-analysis, two approaches that are used increasingly in CER. Additionally, different variations on systematic reviews are being developed to respond to the needs of stakeholders and users (e.g., rapid reviews, evidence maps, scoping reviews; Peterson et al. 2016; Levac, Colquhoun, and O'Brien 2010). Guidance on best practices for conducting systematic reviewers continuously evolves, and researchers should ensure that systematic reviews are conducted consistent with best methodological practices.

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 patients' ability 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 benefit 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 millions of dollars in additional funding.

12: STANDARDS ON RESEARCH DESIGNS USING CLUSTERS

RC-1: Specify whether the study objectives, the interventions, and the primary outcomes pertain to the cluster level or the individual level.
Describe (1) the target population of clusters and individuals to which the study findings will be generalizable, and (2) the clusters to be randomized and the subjects to be enrolled in the trial.

RC-2: Justify the choice of cluster randomization.
Describe the benefits and disadvantages of cluster randomization versus individual-level randomization for the proposed research. Cluster randomization should be substantiated by a sound theoretical and conceptual framework that describes the hypothesized causal pathway (see **CI-1**). Cluster randomization generally is applicable in the following instances:

- An intervention is delivered at the cluster level.
- An intervention changes the physical or social environment.
- An intervention involves group processes.
- An intervention cannot be delivered without a serious risk of contamination.

Logistical considerations can also justify cluster randomization, for example, to reduce costs or to improve participation, adherence, or administrative feasibility.

RC-3: Power and sample size estimates must use appropriate methods to account for the dependence of observations within clusters and the degrees of freedom available at the cluster level.
The methods used to reflect dependence should be clearly described. Sources should be provided for the methods and for the data used to estimate the degree of dependence. Sensitivity analyses that incorporate different degrees of dependence must be reported. For simpler designs, the dependence in the data can be reflected in the intraclass correlation. Dependence can also be reflected in variance components. Other factors that affect the power calculation and should be described include the study design, the magnitude of the hypothesized intervention effect, the prespecified primary analysis, and the desired Type I error rate.

RC-4: Data analyses must account for the dependence of observations within clusters, regardless of its magnitude.
Data analyses must also reflect the degrees of freedom available at the cluster level. Investigators must propose appropriate methods for data analyses, with citations and sufficient detail to reproduce the analyses.

RC-5: Stratified randomization should be used when feasible.
Because cluster randomization trials often involve a limited number of groups or clusters, stratified randomization should be considered and is recommended when feasible. If not feasible, justification should be provided for the use of other methods. The recommended stratification factors are those that are expected to be strongly correlated with the outcome or with the delivery of the intervention, such as baseline value of the outcome variable, cluster size, and geographic area.

Only a limited number of confounders can be addressed through stratification. Other variables, particularly those that characterize the context, should be measured and assessed to document their potential influence on the outcome and understanding of heterogeneity of results.

Rationale for These Standards
Conventional randomized trials allocate individual patients to two or more comparison groups. This is a preferred approach for eliminating systematic differences in the characteristics of the patients in the comparison groups. Randomization of individual patients is ideally suited for studies in which the clinical interventions are standardized and would be expected to have little variation in their delivery to all patients (such as medications). However, many clinical interventions are more complex and depend on decisions, interactions, and processes affected by patients, their providers, and the characteristics of the setting to carry out the intervention (e.g., programs to provide coordinated care in which individual services are sequenced or tailored for individual patients). In these clinical scenarios, both the providers and the setting affect the delivery of clinical care and are an important source of variation in how the services

are provided. When conducting CER of such interventions, it is important to control and/or understand the amount of variation in care delivery within and between clusters to understand the effect of the intervention on patient outcomes.

An approach for controlling variation in the delivery of complex interventions is to change the way in which patients are randomly allocated to receive the clinical interventions being compared. Cluster randomization is an approach in which patients are grouped within units of care delivery (e.g., all patients who receive care from a particular care provider [nurse practitioner, physician, psychologist, physical therapist, etc.], team, or practice). In this approach, the unit of care delivery—rather than the individual patient—is randomized to one of the comparative arms of the study. All patients within that group (the "cluster") are then allocated to that study arm. Cluster randomization has also been advocated as a strategy for evaluating the use of complex interventions in real-world settings in which the investigators have little impact on the intervention's fidelity (Platt et al. 2010).

Cluster randomization has grown in popularity but is not always sufficiently justified. A 2013 systematic review of 73 cluster trials conducted in residential facilities found that only 42 percent provided explicit justification for the cluster design (Diaz-Ordaz et al. 2013). Even in cases where justification is provided, it is sometimes perfunctory and insufficient to support the choice of study design. Guidance on best practices for cluster randomized trials has been provided in published texts (Donner and Klar 2010; Murray 1998) and in recommendations developed by professional groups. The CONSORT Extension for cluster trials published in 2010 provides guidance on how specific objectives and hypotheses should be described (Campbell et al. 2010). These sources emphasize that a cluster design should be used only when justified by the circumstances of the clinical problem being addressed by a study.

Transparency in conceiving, planning, and conducting the study is paramount in helping the scientific community to understand and replicate the study. **RC-1** is a call for transparency and an explicit description of the study objectives, the clinical services being studied, and whether the interventions are targeted at the cluster or the individual level. **RC-2** follows on RC-1 by requiring that the choice of cluster (rather than individual) randomization is justified by the nature of the interventions being examined. Because cluster trials commonly require more participants than an individual randomized trial, proper justification is needed to address the necessity of the research to improve patient outcomes, to document patients' interests in participation, and to ensure protection from unnecessary risks to a larger group of patients.

A challenge in the use of cluster designs is that the clinical outcomes are usually measured at the level of the individual patient, while the unit of randomization is at the cluster level, which requires more complex statistical methods (**RC-3** and **RC-4**). When using the patient as the unit of analysis, the analytic approach must account for the clustering and the consequent correlations among the patients in each cluster. In other words, cluster randomization threatens the assumption that all patients are independent from each other. It also results in a loss of statistical power compared with an approach in which randomization was performed at the level of each individual patient.

RC-3 emphasizes the importance of realistic estimates of statistical power for cluster designs. In particular, researchers should avoid using unrealistically low estimates of the degree of similarity within clusters (usually represented by the intraclass correlation coefficient). Prior studies have found that the intraclass correlation can be unexpectedly large (Koepsell 1998; Verma and Le 1996). When making power estimates for a planned cluster-based study, it is prudent to use a sufficiently large estimate of intra-cluster correlation.

RC-4 addresses the need for adjustments in the analysis, if there is substantial variation in the number of individuals enrolled in the individual clusters after study completion. When some clusters have small sample sizes, the effective degrees of freedom should be reduced to reflect that these clusters cannot meaningfully contribute to the analysis (Murray 1998).

Finally, stratified randomization should be used when feasible (**RC-5**). Cluster randomized trials often involve a limited number of clusters, which may reduce the likelihood that randomization will produce similar distributions of potential confounders across the clusters. In addition, because only a limited set of confounders can be addressed through stratification, other variables—particularly those that characterize the context of the intervention—should be measured and their potential influence on the estimates of the interventions' effects assessed and documented in study reports.

13: STANDARDS FOR STUDIES OF COMPLEX INTERVENTIONS

SCI-1: Fully describe the intervention and comparator and define their core functions.

Describe the intervention and comparator under study and clearly define aspects related to core functions and forms. Core functions refer to the intended purpose(s) of the interventions. The form of the interventions includes the intended modes of delivery, providers involved, materials or tools required, dose, and frequency/intensity. The description should also explicitly indicate to whom the intervention is aimed (e.g., patient, provider, hospital, health system).

SCI-2: Specify the hypothesized causal pathways and their theoretical basis.

Clearly describe the hypothesized causal pathways by which the proposed complex intervention generates change (see **CI-1**). This description should depict how each intervention function generates the hypothesized effects on the prespecified patient outcome(s). Include in the causal model key contextual factors that may influence the impact of the intervention so that their hypothesized relationships are made explicit. Describe the theoretical and/or empirical bases underlying the proposed interventions and their hypothesized effects.

SCI-3: Specify how adaptations to the form of the intervention and comparator will be allowed and recorded.

Specify any allowable adaptations in form and describe how planned and unplanned adaptations will be managed, measured/documented, and reported over time. Any planned adaptations should have a clear rationale; be supported by theory, evidence, or experience; and maintain fidelity to the core functions of the intervention. Upon conclusion of the study, researchers should provide guidance on allowable adaptations or unproductive adaptations (i.e., adaptations that may reduce the effectiveness of an intervention).

SCI-4: Plan and describe a process evaluation.

Describe plans to conduct a process evaluation (i.e., to assess whether the intervention was implemented as planned and to test and refine the hypothesized causal pathways). Process evaluations should measure/document, analyze, and report the fidelity of the delivery of the intervention (i.e., planned and unplanned adaptations); the quantity or dose of the intervention actually delivered; whether the intended population(s) received the delivered intervention (i.e., reach); the mechanisms of action (e.g., mediators, intermediate outcomes); and important contextual factors (e.g., moderators), taking into account the levels at which the intervention is aimed (e.g., patient, provider, hospital).

Researchers should select a combination of methods appropriate to the process questions identified and describe the timing and sources of data collection. These plans should include appropriate quantitative, qualitative, and/or mixed methods that account for the intervention functions as defined by the causal pathway.

Describe the plans to integrate process and outcome data in advance of intervention delivery to determine whether and how outcomes and effects are influenced by implementation or contextual moderators. Explain how the results of the process evaluation will be used to draw inferences about both the effectiveness (i.e., patient outcomes) and the processes of care (i.e., process outcomes).

SCI-5: Select patient outcomes informed by the causal pathway.

Select valid and reliable patient outcome measures that are explicitly affected by the hypothesized causal pathway and the theoretical and/or empirical basis for the intervention. If the study does not measure a patient outcome, researchers must provide strong evidence that supports the linkage between the measured outcome and unmeasured patient outcome. The outcome measures should assess the intervention across a range of domains that sufficiently permit assessment of how the intervention affects patients. In determining the length of follow-up, assumptions about the rate and pattern of change expected in the outcome measures should be clear.

Rationale for These Standards

Many healthcare interventions require specific involvement and behaviors by patients, caregivers, and healthcare providers. A complex intervention is a multicomponent intervention that may act independently or interdependently to change patient outcomes (Craig et al. 2013). Examples include various nonpharmacological treatments, behavioral interventions, lifestyle modifications, and reorganization of specific aspects of the delivery system. In comparative effectiveness research, the intervention, the comparator, or both may be complex. In general, a complex intervention usually has one or more of the following characteristics (Craig et al. 2013; Guise et al. 2017):

- Multiple components that interact
- Specified behaviors and activities carried out by individuals (e.g., healthcare staff, providers, patients, caregivers)

- Multiple entities or levels targeted by the intervention
- Contextual factors associated with variation in outcomes.

To facilitate transparency and replicability of research findings, complex interventions must be fully and specifically described. Investigators should describe the essential functions of an intervention (**SC-1**), which should be supported by theory and/or evidence and reflected in their causal model **(SCI-2**). An intervention's function refers to the key mechanisms and processes that achieve an intended purpose (Byng et al. 2008; Hawe 2015; Hawe, Shiell, and Riley 2004). Defining the core functions of a complex intervention facilitates comparability, replicability, and adaptation by clarifying the specific underlying processes that are hypothesized to be responsible for the change in patient outcomes.

SCI-1 also requires a description of the form of the intervention, which includes the components and characteristics that are required to achieve the intended functions. Examples of this description may include modes of delivery, providers involved, materials or tools required, dose, and frequency/intensity of the intervention (Hoffmann et al. 2014; Möhler, Köpke, and Meyer 2015). A clear description of form is fundamental to the interpretation of study results and reliable implementation of interventions that have positive findings. While fidelity to function is key, researchers must specify what aspects of form should be standardized within their study (**SCI-3**), as it may influence the intervention's effectiveness and replicability. The description should also explicitly indicate to whom the intervention is aimed (e.g., patient, provider, hospital, health system). Researchers may consult the TIDieR and/or CReDECI-2 checklists for guidance and use of consistent terminology to permit comparably across studies (Hoffmann et al. 2014; Möhler, Köpke, and Meyer 2015).

SCI-2 requires investigators to outline the hypothesized causal pathways by which the proposed intervention generates change in patient outcomes. Complex interventions often involve multiple causal pathways with mediators and moderators that have either direct or indirect impacts on patient outcomes (Guise et al. 2017). Key contextual factors that may influence the impact of the intervention should be included in the causal model so that their hypothesized relationships are made explicit. Contextual factors may interact or influence the intervention, thereby diminishing or enhancing its effectiveness. The theoretical and/or empirical bases underlying the proposed interventions and their hypothesized effects should be described and the strength of this evidence should be made explicit. The causal pathway may be depicted visually to illustrate the intended interactions between the intervention ingredients and outcomes (Moore et al. 2015).

A well-described or depicted causal pathway, along with possible mechanisms of action and relevant contextual factors, informs selection of relevant patient outcomes and provides the basis for a well justified data analysis plan (Craig et al. 2013). Making these relationships explicit may also inform the investigator's choice of comparator. These descriptions add transparency ahead of the study implementation and can aid in improving scientific rigor, study implementation, and replicability of the results once the study is complete.

SCI-3 requires investigators to think through the possible adaptations to the design of the intervention (i.e., through a selection of suitable and equivalent forms) before study initiation, as opposed to making arbitrary or ad hoc decisions while the study is underway. Many contextual factors, such as time, resources, training, organizational context, language, and culture, can challenge the faithful implementation of a complex intervention (Greenhalgh et al. 2004). Adaptations are often made in response to these local circumstances. Planned adaptations must be explicitly outlined; should have a clear rationale; be supported by theory, evidence, or experience; and maintain fidelity to the core functions of the intervention (Bauman, Stein, and Ireys 1991). Existing tools and frameworks may help researchers think through their allowable adaptations while preserving the intervention's causal model (Bauman, Stein, and Ireys 1991; Stirman et al. 2013).

Researchers must explicitly outline how planned and unplanned adaptations will be managed, measured/documented, and reported over time. Unplanned, observed adaptations that were not prespecified should also be documented, including the rationale, setting, and frequency of those adaptations.

Upon conclusion of the study, researchers should provide guidance on allowable adaptations or unproductive adaptations (i.e., adaptations that may reduce the effectiveness of an intervention). It may not be possible to infer the marginal impact of adaptations to a complex intervention. Existing guidance may offer ways to categorize adaptations to help communicate them to a broader audience (Stirman et al. 2013).

Investigators should also plan and conduct a process evaluation to determine whether and how the intervention—as it was delivered during the study—achieved its intended effects (**SC-4**). Process evaluations are key to transparency and replicability of study results and may help explain discrepancies between expected and observed outcomes.

Data collection should be informed by the causal pathway and a theoretical understanding of how the intervention causes change (Moore et al. 2015). The process-related questions should relate to links in the causal chain, rely on an established framework, and be defined and planned a priori. Data should be collected on the process (how delivery is achieved), fidelity, adaptations, dose, and reach. Mechanisms of action should be measured or documented to evaluate the hypothesized causal pathways and better understand how an intervention affected patients. Process evaluations also permit researchers to describe and report on the context within which the intervention was delivered, including any external barriers or facilitators that influenced its delivery. Furthermore, participant burden must be considered when developing data collection plans.

Many process evaluations use a combination of quantitative, qualitative, and/or mixed methods approaches (Moore et al. 2015; Raine et al. 2016). Quantitative methods may be suitable for capturing descriptive information on fidelity, dose, and reach as well as measuring key process variables and testing hypothesized mechanisms of impact (mediational analysis) and contextual moderators (Emsley, Dunn, and White 2010; Moore et al. 2015). Some process questions may be more appropriately answered by qualitative approaches, such as whether changes in implementation emerge during intervention delivery and by what means, or how recipients at multiple levels experience the intervention (Bonell et al. 2012; Kane et al. 2014). Qualitative data may also generate new theories to be tested in follow-up studies. To the extent possible, researchers should collect data at multiple time points to assess change in implementation or account for contextual factors. Researchers should specify in advance whether process evaluations will be done independently or concurrently with outcomes evaluation and adhere to **SCI-3** if the process evaluation results will be used to inform study implementation.

While process outcomes may be informative for causal inference, they are often insufficient on their own to assess an intervention's effectiveness. Researchers should therefore ensure that they have also selected appropriate patient outcomes that are explicitly affected by the hypothesized causal pathway (**SC-5**). Additional justification is required in the (rare) cases that a study does not also include measurement of patient outcomes that are hypothesized to be influenced by the complex intervention.

Several organizations have developed guidelines and best practices for developing, evaluating, and reporting complex interventions, which investigators may consult for additional guidance (Boutron et al. 2017; Coly and Parry 2017; Craig et al. 2013; Hoffmann et al. 2014; Moore et al. 2015).

SECTION IV: ADVANCING UNDERSTANDING AND APPROPRIATE USE OF METHODS FOR PCOR

Good research practices are a required foundation for high-quality PCOR. One of the most important components of good practices is a commitment to transparency, which enables other researchers to assess the reproducibility and validity of findings. Many of the PCORI Methodology Standards promote transparency by requiring detailed protocols before beginning the research and compliance with guidelines when results are reported. These requirements help PCORI and others judge the quality and relevance of the research and help protect against practices that can distort or misrepresent research results, such as selective reporting (Chan et al. 2014; Glasziou et al. 2014).

PCORI uses a comprehensive, coordinated approach to promote the wide use of its methodology standards. Strategies to support adoption include engaging a broad range of stakeholders who use or might use the standards; collaborating with other organizations and initiatives to strengthen research practices and facilitate use of the standards; using reporting and surveillance mechanisms; and offering multiple resources, including in-person and web-based training opportunities. Other initiatives include outreach to both professional and public audiences to promote use and adoption of best practices for PCOR.

PCORI has a commitment to evaluate and update the guidance that it provides to the research community. In its ongoing work, PCORI's Methodology Committee follows a process to update, refine, and expand the scope of its methodological guidance in areas where minimum standards can strengthen PCOR questions and approaches. The Methodology Committee is currently undertaking work to develop methodology standards in a number of areas, including individual participant data meta-analysis, data quality, and qualitative and mixed methods. Consistent with this work and advances in research methodology, future editions of the Methodology Report and Standards will provide updated methodological guidance for PCOR to support the generation of high-quality and relevant evidence that patients, clinicians, and other stakeholders need to make informed health decisions.

APPENDIX A: PCORI METHODOLOGY STANDARDS

1: STANDARDS FOR FORMULATING RESEARCH QUESTIONS

RQ-1: Identify gaps in evidence.
Gaps in the evidence identified in current systematic reviews should be used to support the need for a proposed study. If a systematic review is not available, one should be performed using accepted standards in the field (see **SR-1**), or a strong rationale should be presented for proceeding without a systematic review. If the proposed evidence gap is not based on a systematic review, the methods used to review the literature should be explained and justified.

RQ-2: Develop a formal study protocol.
Researchers should develop a formal protocol that provides the plan for conducting the research. The protocol should specify the research objectives, study design, exposures and outcomes, and analytical methods in sufficient detail to support appropriate interpretation and reporting of results. Protocols should be submitted to the appropriate registry (e.g., clinicaltrials.gov), and all amendments and modifications (e.g., changes in analytic strategy, changes in outcomes) should be documented.

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(s) for whom the health decision is pertinent, and (3) how study results will inform the health decision.

RQ-4: Identify and assess participant subgroups.
When designing studies, researchers should identify participant subgroups, explain why they are of interest, and specify whether subgroups will be used to test a hypothesis or to conduct an exploratory analysis, preferably based on prior data. A study should have adequate precision and power if conclusions specific to these subgroups will be reported.

RQ-5: Select appropriate interventions and comparators.
The interventions and comparators should correspond to the actual healthcare options for patients, providers, and caregivers who would face the healthcare decision. The decision should be of critical importance to the relevant decision makers, and one for which there is a compelling need for additional evidence about the benefits and harms associated with the different options. Researchers should fully describe what the comparators are and why they were selected, describing how the chosen comparators represent appropriate interventions in the context of the relevant causal model (**CI-1**), reduce the potential for biases, and allow direct comparisons. Generally, usual care or nonuse comparator groups should be avoided unless they represent legitimate and coherent clinical options.

RQ-6: Measure outcomes that people who represent the population of interest notice and care about.
Identify and include outcomes that the population of interest notices and cares about (e.g., survival, functioning, 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 that reflect both beneficial and harmful effects, based on input from patient informants and people representative of the population of interest.

2: STANDARDS ASSOCIATED WITH PATIENT CENTEREDNESS

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

Include individuals affected by the condition and, as relevant, their surrogates and/or caregivers. Other relevant stakeholders may include, but are not limited to, clinicians, purchasers, payers, industry, hospitals, health systems, policy makers, and training institutions. These stakeholders may be end users of the research or be involved in healthcare decision making.

As applicable, researchers should describe how stakeholders will be identified, recruited, and retained and the research processes in which they will be engaged. Researchers should provide a justification in proposals and study reports if stakeholder engagement is not appropriate in any of these processes.

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 the following:

- The plan to ensure representativeness of participants
- How participants are identified, selected, recruited, enrolled, and retained in the study to reduce or address the potential impact of selection bias
- Efforts employed to maximize adherence to agreed-on enrollment practices
- 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 (e.g., individuals historically underrepresented in healthcare research, such as those with multiple disease conditions, low literacy, low socioeconomic status, or poor healthcare access; racial and ethnic minority groups; people living in rural areas), then specify plans to address population-specific 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 for outcomes of interest.

To measure outcomes of interest identified as patient centered and relevant to decision makers (see **RQ-6**) and for which patients or people at risk of a condition are the best source of information, the study should employ patient-reported outcome (PRO) measures and/or standardized questionnaires with appropriate measurement characteristics for the population being studied. When selecting PRO measures for inclusion in a study, researchers, in collaboration with patient and other stakeholder partners, should consider (1) the concept(s) underlying each PRO measure (e.g., symptom, 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 its 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 as well as the translation process if the measure is to be used in multiple languages. If these measurement properties are not known, a plan to establish the properties must be provided. Caregiver reports may be appropriate if the patient cannot self-report the outcomes of interest.

PC-4: Support the dissemination and implementation of study results.

All study results must be made publicly available. To ensure study objectives and results are understandable and actionable by as many people as possible, they should be presented in lay language summaries. For study results that are appropriate for dissemination and implementation, involve patients and other relevant stakeholders in (1) planning for dissemination from the start of the research study, (2) creating a dissemination plan for the study that indicates clinical implications, (3) working with patients or organizations to report results in a manner understandable to and usable by each target audience, and (4) identifying successful strategies for the adoption and distribution of study findings to targeted patient and clinical audiences.

3: STANDARDS FOR DATA INTEGRITY AND RIGOROUS ANALYSES

IR-1: A priori, specify plans for quantitative data analysis that correspond to major aims.
Before analysis is undertaken, researchers should describe the analytic approaches that will be used to address the major research aims. These include definitions of key exposures, outcomes, and co-variates. As applicable, study protocols should identify patient subgroups of interest, plans (if any) for how new subgroups of interest will be identified, and how analysis plans may be adapted based on changing needs and scientific advances. Researchers should also specify plans for handling missing data and assessing underlying assumptions, operational definitions, and the robustness of their findings (e.g., sensitivity analyses).

IR-2: Assess data source adequacy.
In selecting data sources and planning for data collection, researchers should ensure the robust capture of exposures or interventions, outcomes, and relevant covariates. Measurement properties of exposures and outcomes should be considered, and properties of important covariates should be taken into account when statistically adjusting for covariates or confounding factors.

IR-3: Describe data linkage plans, if applicable.
For studies that link patient data from two or more sources (including registries, data networks, and others), describe (1) the data sources and/or the linked data set in terms of 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(s).

IR-4: Document validated scales and tests.
Studies should include documentation of the names of the scales and tests selected, reference(s), characteristics of the scale, and psychometric properties.

IR-5: 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 lists 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), STROBE (for observational studies), and SRQR and/or COREQ (for studies that use qualitative research). Researchers should register their studies with the appropriate registry (e.g., clinicaltrials.gov for clinical studies or observational outcomes studies) and provide complete and accurate responses to the information requested (e.g., enter the required and optional data elements for clinicaltrials.gov).

IR-6: Masking should be used when feasible.
Masking (also known as blinding) of research staff should be implemented, especially in situations for which study participant and investigator masking are not feasible. When masking is not feasible, the impact of the lack of masking on the results should be discussed.

IR-7: In the study protocol, specify a data management plan that addresses, at a minimum, the following elements: collecting data, organizing data, handling data, describing data, preserving data, and sharing data.

Data management is a critical phase in clinical research that contributes to the generation of high-quality, reliable, and statistically sound data from clinical trials and observational studies. The underlying motivation for good data management practice is to ensure that the data are accessible, sustainable, and reproducible, both for future investigators and for the original research team. This standard applies to both the quantitative and the qualitative data collected in a study.

A data management plan (DMP) is a document that describes what data will be generated by a research study, how the

data will be managed and stored, who will have access to the data, what documentation and metadata will be created with the data, how the data will be preserved, and how the data will be shared in support of future scientific inquiries. DMPs are distinct from statistical analysis plans, which describe the planned statistical analyses associated with the study (e.g., statistical tests to be used to analyze the data, how missing data will be accounted for in the analysis).

To ensure quality control, the study investigators should self-monitor their data management procedures. This includes conducting checks to ensure manually entered subject numbers conform to study-defined site/subject number format rules and conducting real-time review of data to verify their accuracy and validity.

DMPs should include language that, at a minimum, addresses each of the following considerations:

- Collecting data: Based on the hypotheses and sampling plan, describe what data will be generated and how they will be collected. Provide descriptive documentation of the data collection rationale and methods, and any relevant contextual information.
- Organizing data: Decide and document how data will be organized within a file, what file formats will be used, and what types of data products will be generated.
- Handling data: Describe and document who is responsible for managing the data, how version control will be managed, what the data handling rules are, what the method and frequency for backing up the data will be, and how confidentiality and personal privacy will be protected.
- Describing data: Describe how a data dictionary and metadata record will be produced (i.e., metadata standard and tools that will be used).
- Storing and preserving data: Implement a data storage and preservation plan that ensures that both the raw data and the analytic files can be recovered in the event of file loss. Document the data storage and preservation plan, including the approach to data recovery (e.g., routinely storing data in different locations).
- Maintaining data: Develop a plan to maintain the data in a data repository.
- Sharing data: Develop a plan to share data with the project team, with other collaborators, and with the broader scientific community.

Consistent with the Guideline for Good Clinical Practice, the investigator/institution should maintain adequate and accurate source documents, including the DMP. The DMP should be attributable, contemporaneous, original, accurate, and complete. Changes to the DMP should be traceable, should not obscure the original entry, and should be explained if necessary (e.g., via an audit trail).

4: STANDARDS FOR PREVENTING AND HANDLING MISSING DATA

MD-1: Describe methods to prevent and monitor missing data.

Investigators should explicitly state potential reasons that study data may be missing. Missing data can occur from patient dropout, nonresponse, data collection problems, incomplete data sources, and/or administrative issues. As relevant, the protocol should include the anticipated amount of and reasons for missing data, plans to prevent missing data, and plans to follow up with study participants. The study protocol should contain a section that addresses steps taken in study design and conduct to monitor and limit the impact of missing data. This standard applies to all study designs for any type of research question.

MD-2: Use valid statistical methods to deal with missing data that properly account for statistical uncertainty owing to missingness.

Valid statistical methods for handling missing data should be prespecified in study protocols. The analysis should explore the reasons for missing data and assess the plausibility of the assumptions associated with the statistical methods. The potential impact of missing data on the results and limitations of the approaches used to handle the missing data should be discussed.

Estimates of treatment effects or measures of association should be based on statistical inference procedures that account for statistical uncertainty attributable to missing data. Methods used for imputing missing data should produce valid confidence intervals and permit unbiased inferences based on statistical hypothesis tests. Bayesian methods, multiple imputation, and various likelihood-based methods are valid statistical methods for dealing with missing data. Single imputation methods, such as last observation carried forward, baseline observation carried forward, and mean value imputation, are discouraged as the primary approach for handling missing data in the analysis. If single imputation–

based methods are used, investigators must provide a compelling scientific rationale for why the method is appropriate. This standard applies to all study designs for any type of research question.

MD-3: 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 participation in all or only some study activities. Investigators should attempt to continue to collect information on key outcomes for participants unless consent is withdrawn. All participants included in the study should be accounted for in study reports, regardless of whether they are included in the analyses. Any planned reasons for excluding participants from analyses should be described and justified. In addition, missing data owing to other mechanisms (such as nonresponse and data entry/collection) should be documented and addressed in the analyses.

MD-4: Examine sensitivity of inferences to missing data methods and assumptions, and incorporate it into the 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, 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, when possible, the abstract of any reports.

5: STANDARDS FOR HETEROGENEITY OF TREATMENT EFFECTS (HTE)

HT-1: State the goals of HTE analyses, including hypotheses and the supporting evidence base.
State the inferential goal of each HTE analysis and explain how it relates to the research topic. Specify whether the HTE analysis is hypothesis driven (sometimes denoted as confirmatory) or hypothesis generating (sometimes denoted as exploratory). Hypothesis-driven HTE analyses should be prespecified based on prior evidence (described clearly in the study protocol and study reports) and supported by a clear statement of the hypotheses the study will evaluate, including how subgroups will be defined (e.g., by multivariate score, by stratification), outcome measures, and the direction of the expected treatment effects.

HT-2: For all HTE analyses, provide an analysis plan, including the use of appropriate statistical methods.
The study protocol should unambiguously prespecify planned HTE analyses. Appropriate methods 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. Appropriate methods should be used to account for the consequences of multiple comparisons; 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).

HT-3: Report all prespecified HTE analyses and, at minimum, the number of post hoc HTE analyses, including all subgroups and outcomes analyzed.
Both protocols and study reports must report the exact procedures used to assess HTE, including data mining or any automatic regression approaches. HTE analyses should clearly report the procedures by which subgroups were defined and the effective number of subgroups and outcomes examined. Within each subgroup level, studies should present the treatment effect estimates and measures of variability. Prespecified HTE analyses (hypothesis driven) should be clearly distinguished from post hoc HTE analyses (hypothesis generating). Statistical power should be calculated and reported for prespecified (hypothesis-driven) analyses.

6: STANDARDS FOR DATA REGISTRIES

DR-1: Requirements for the design of registries

Registries established to conduct patient-centered outcomes research (PCOR) must have the following characteristics:

A. Registry Purpose and Protocol. The purpose of the registry should be clearly defined to guide the design of key registry features, including, but not limited to, the target population, the research question(s) to be addressed, the data source used, the data elements collected, data-sharing policies, and the stakeholders involved in developing and using the registry. Participants and other key stakeholders should be engaged in registry design and protocol development. Registries should aim to be user oriented in design and function.

B. Data Safety and Security. Registry custodians should comply with institutional review board (IRB) human subjects protection requirements, the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule, and all other applicable local, state, and national laws. Registries should provide information that describes the type of data collection (primary or secondary source data), data use agreements, informed consent documents, data security protections, plans to maintain data protection if the registry ends, and approaches to protect privacy, including risk of and/or process for re-identification of participants, especially for medical or claims records.

C. Data Elements and Quality. Standardized data element definitions and/or data dictionaries should be used whenever possible. When creating a new registry, published literature should be reviewed to identify existing, widely used definitions of outcomes, exposures, and confounders before new definitions are drafted.

When collecting primary data, conduct multistakeholder engagement with potential participants and data users to prioritize data collection needs. When participants support their face validity, use validated instruments or PRO measures when available. If secondary data sources (e.g., electronic medical records, claims data) are used, describe the original purpose of the secondary data and verify the accuracy and completeness of the data, as well as the approach to and validity of the linkages performed between the primary and secondary sources.

The specifics of the quality assurance plan will depend on the type of data (primary or secondary) collected by the registry. In general, the plan should address (1) structured training tools for data abstractors/curators; (2) the 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 (where feasible and appropriate), and validation statistics focused on data quality for the key exposure and outcome variables and key covariates. A risk-based approach to quality assurance, focused on variables of greatest importance, is advisable.

D. Confounding. During the planning phase, registries should identify important potential confounders pertinent to the purpose and scope of the research. During the analysis phase, they should collect reasonably sufficient data on these potential confounders to facilitate the use of appropriate statistical techniques. When conducting analyses, refer to the PCORI Methodology Standards for Data Integrity and Rigorous Analyses and Standards for Causal Inference Methods.

E. Systematic Participant Recruitment and Enrollment. Develop a sampling plan of the target population and identify recruitment strategies for participants that minimize the impact of selection bias. Participants should be enrolled systematically, with similar procedures implemented at all participating sites and for each intervention of interest. Confirm adherence to agreed-on enrollment practices.

F. Participant Follow-Up. The objective(s) of the registry should determine the type, extent, and length of participant follow-up.

Describe the frequency with which follow-up measures will be ascertained, consider linkage with other data sources (e.g., the National Death Index) to enhance long-term follow-up, and identify the date of last contact with the participant in existing registries, where appropriate. Ensure that the participants are followed in as unbiased a manner as possible, using similar procedures at all participating sites.

Monitor loss to follow-up to ensure best efforts are used to achieve follow-up time that is adequate to address the main objective. At the outset of the registry, develop a retention plan that documents when a participant will be considered lost to follow-up and which actions will be taken to minimize the loss of pertinent data. Retention

efforts should be developed with stakeholders to ensure that the efforts are suitable for the target population and that anticipated challenges are addressed appropriately.

DR-2: Documentation and reporting requirements of registry materials, characteristics, and bias

Clearly describe, document with full citations where appropriate, and make publicly available registry materials, including, but not limited to, registry protocols, data-sharing policies, operational definitions of data elements, survey instruments used, and PROs captured. Modifications to any documents or data collection instruments should be clearly described and made available for registry users and participants. Characteristics of the registry participants should be described. To help assess potential selection biases, identify how the participants may differ from the target population. Document the loss to follow-up and describe the impact on the results, using sensitivity analyses (prespecified where possible) to quantify possible biases. Report the extent of bias clearly to stakeholders who may want to use the registry resource.

DR-3: Adapting established registries for PCOR

Previously established registries that intend to support new clinical research may not have been informed by all applicable methodology standards. When new research will use such registries, investigators should engage key stakeholders, including registry participants, to assess the feasibility of using the registry for new research and to ensure the following:

- Informed consent documents are appropriately tailored to participant needs, characteristics, and conditions.
- Data elements are meaningful and useful to researchers and participants.
- Recruitment and retention strategies are feasible and effective.
- Registry policies are patient centered and the use of registry data is transparent to participants.
- Dissemination practices are appropriate and effective at reaching the communities from which the data are collected.
- Opportunities for bidirectional benefit exist between participants and researchers.
- Registry materials, described in **DR-2**, and informed consent forms are publicly available in accessible formats.

DR-4: Documentation requirements when using registry data

Researchers planning PCOR studies that rely on registries must ensure that these registries meet the requirements contained in **DR-1** and **DR-2** and must document each required feature of each registry to be used (e.g., in an appendix to the funding application or study protocol). Deviations from the requirements in **DR-1** and **DR-2** should be well documented and limitations of research related to the deviations from requirements should be addressed when reporting study findings.

7: STANDARDS FOR DATA NETWORKS AS RESEARCH-FACILITATING STRUCTURES

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

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

A. **Data Integration Strategy.** 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 (including code and process documentation) 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-identifying data and apply algorithms to ensure that the desired level of confidentiality is achieved to meet the particular PCOR application's need. Data custodians should ensure that the data privacy/consents of the original data source cover the intended data usage through the data network. Privacy protections, including which data will be released and how breaches will be addressed, should be specified in the data use agreement. The physical security of the data and data platforms should be considered and addressed as well.

C. **Identity Management and Authentication of Individual Researchers.** Develop reliable processes to verify and authenticate the credentials of researchers who are granted access to a distributed research network.

D. **IP Policies.** A research network should develop policies for handling and disseminating 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 that uses the network. Guidelines should balance (1) minimizing impediments to innovation in research processes and (2) making the research results widely accessible, particularly to the people who need them the most.

E. **Standardized Terminology Encoding of Data Content.** The data content should be represented with a clearly specified standardized terminology system to ensure that its meaning is unambiguously and consistently understood by parties using the data.

F. **Metadata Annotation of Data Content.** Semantic and administrative aspects of data content 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 organized into a standard structure that establishes common definitions and shows close or distant associations among variables. A common data model specifies necessary data items that need to be collected and shared across participating institutes, clearly represents the 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 that rely on data networks must ensure that these networks meet the requirements contained in **DN-1**, and they must document the current maintenance status of the data network (e.g., currency of the data, level of data curation). Because different studies are expected to have different dependencies on various components of the data network, researchers should assess the appropriateness of the data in the network for a specific research study through the following activities:

A. **Data content and conformance.** Document what is actually needed for the research question and compare that to the sources in the network. Identify which data are best represented by the network's data sources and how they are included in the study. Ensure that the representations and values of the data to be used from the network are sufficient to address the research question.

B. **Data quality.** Assess the data quality for the data sources that will be used. It is especially important to assess data completeness and plausibility. Where data are incomplete, identify and assess potential biases for completeness and consider alternate sources. Assess plausibility by reviewing data value distributions and comparing additional data sources that would have expected concordance with the selected sources. Determine whether the data sources are of sufficient quality to be included in the analysis.

C. **Sensitivity analyses.** After the initial analysis is completed, perform sensitivity analyses on the data sources to test whether possible variations in data characteristics would affect the conclusions of the analysis. Specifically, measure the sensitivity of the conclusions to the following:

- Completeness and correctness of the data in the data network
- Availability of data sources that are most likely at risk of exclusion
- Temporal dependence of the data
- Operational definitions and decisions made to implement analysis

The results of these assessments should be documented and included with any findings from research studies that use the data networks.

8: STANDARDS FOR CAUSAL INFERENCE METHODS

CI-I: Specify the causal model underlying the research question *(cross-cutting standard, applies to all PCOR/CER studies)*. Researchers should describe the causal model relevant to the research question, which should be informed by the PICOTS framework: populations, interventions, comparators, outcomes, timing, and settings. The causal model represents the key variables; the known or hypothesized relationships among them, including the potential mechanisms of effect; and the conditions under which the hypotheses are to be tested. Researchers should use the causal model to determine whether and how the study can handle bias and confounding and the extent to which valid estimates of the effects of an intervention can be generated given the particular hypothesis, study design, analytical methods, and data source(s).

CI-2: Define and appropriately characterize the analysis population used to generate effect estimates. Researchers should specify the eligibility criteria for inclusion in the study population and analysis. In prospective studies, decisions about which patients are included in an analysis should be based on information available at each patient's time of study entry. In retrospective studies, these decisions should be based on information from a defined period before the start of exposure. For time-varying treatment or exposure regimes, specific time points should be clearly specified; relevant variables measured at baseline and up to, but not beyond, those time points should be used as population descriptors. 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), address selection bias that may be introduced by excluding patients, and assess the potential impact on the validity of the results.

CI-3: Define with the appropriate precision the timing of the outcome assessment relative to the initiation and duration of exposure. To reduce potential sources of bias that arise from inappropriate study design choices (e.g., immortal time bias), researchers must precisely define, to the extent possible, the timing of the outcome assessment relative to the initiation and duration of the exposure.

CI-4: Measure potential confounders before the start of exposure and report data on potential confounders with study results. In general, variables used in confounding adjustment (either in the design or analysis) should be ascertained and\ measured before the first exposure to the intervention(s) 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 adjust for measured confounding, researchers should consider and report how propensity scores will be created (high dimensional propensity score versus a priori clinical variables) and which balancing method will be used (e.g., matching, weighting, stratifying). Researchers should assess and report 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. When an instrumental variable (IV) approach is used (most often to address unmeasured confounding), empirical evidence should be presented that describes how the variable chosen as an IV satisfies the three key properties of a valid instrument: (1) the IV influences the choice of 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, decisional thresholds, and statistical properties of those adaptations.

The adaptive clinical trial design must be prospectively planned and the design must be clearly documented in the study protocol before trial enrollment begins, including at a minimum the following:

- All potential adaptations, including timing
- Interim trial findings that will be used in determining each adaptation
- Statistical models and decisional thresholds to be used
- Planned analyses of the trial endpoint(s)

The description of the design should be sufficiently detailed that it could be implemented based on the description of procedures. This specification should include a statistical analysis plan in which all necessary detail is provided regarding planned interim and final analyses.

Additionally, the statistical properties of 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, 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.

AT-2: Specify the 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 about how the prior distribution was determined and if an informative or noninformative 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. Computational issues should be addressed, including describing the choice of software and creating, testing, and validating that software. Software used for Bayesian calculations during the trial design, execution, and final analysis must be functionally equivalent. When feasible, software or other computing packages should be made available to relevant stakeholders for evaluation and validation.

AT-3: Ensure that clinical trial infrastructure is adequate to support planned adaptation(s) and independent interim analyses.

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 because this information provides data regarding treatment success.

AT-4: When reporting adaptive randomized clinical trials, use the CONSORT statement, with modifications.

The following sections of the 2010 CONSORT statement can be used to report key dimensions of adaptation:

- Adapting randomization probabilities (sections 8b and 13a)
- Dropping or adding study arms (sections 7b and 13a)
- Interim stopping for futility and superiority or adverse outcomes (sections 7b and 14b)
- Reestimating sample size (sections 7a and 7b)

- Transitioning stages (e.g., seamless Phase II/III designs; sections 3a, 7a, 7b, and 16)
- Modifying inclusion and exclusion criteria (sections 4a and 13a)

CONSORT sections 16, 20, and 21 provide additional guidance on reporting aspects of an adaptive trial.

All possible adaptations included in the prospective design, even if they did not occur, should be included in the study reports.

10: STANDARDS FOR STUDIES OF MEDICAL TESTS

(Formerly Standards for Studies of Diagnostic Tests)

MT-1: Specify the clinical context and key elements of the medical test.

Evaluation of tests used to inform medical decision making (e.g., diagnostic tests, prognostic tests, predictive tests) should specify each of the following items and provide justification for 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 choice of comparator (e.g., another test, no test) and goal of the comparison; (3) the technical specifications of the test(s) as implemented in the study; (4) the approach to test interpretation; (5) the sources and process for obtaining reference standard information, when applicable; (6) the procedures for obtaining follow-up information and determining patient outcomes, when applicable; and (7) the clinical pathways involving the test and the anticipated implications of test use on downstream processes of care and patient outcomes. These items ought to be specified for all types of tests used for medical decision making and 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.

MT-2: Assess the effect of factors known to affect performance and outcomes.

Studies of tests used to inform medical decision making should include an assessment of the effect of important factors known to affect test performance and outcomes, including, but not limited to, the threshold for declaring a "positive" test result, the technical characteristics of the test, test materials (e.g., collecting, preparing, and handling samples), operator dependence (e.g., lab quality, interpretation requirements), and the care setting.

MT-3: Focus studies of medical tests on patient-centered outcomes, using rigorous study designs with a preference for randomized controlled trials.

A prospective randomized design should be used when possible to assess the diagnostic, prognostic, predictive, and/or therapeutic outcomes of testing. If a nonrandomized design is proposed, a rationale for using an observational study (or modeling and simulation) should be provided, and efforts to minimize confounding documented.

11: STANDARDS FOR SYSTEMATIC REVIEWS

SR-1: Adhere to National Academy of Medicine (NAM) standards for systematic reviews of comparative clinical effectiveness research, as appropriate.

Systematic reviews, which critique and synthesize the existing literature, can also identify evidence gaps and inform decisions of how to address these gaps. Existing standards for systematic reviews developed by credible authorities, such as the Cochrane Collaboration and the Agency for Healthcare Research and Quality, vary somewhat in their recommended approaches. The PCORI Methodology Committee endorses the standards issued by NAM in 2011 but recognizes both the importance of conducting systematic reviews consistent with updates to best methodological practices and that there can be flexibility in the application of some standards without compromising the validity of the review, including the following:

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

Researchers should describe and justify any departures from the 2011 NAM standards (e.g., why a particular requirement does not apply to the systematic review).

12: STANDARDS ON RESEARCH DESIGNS USING CLUSTERS

RC-1: Specify whether the study objectives, the interventions, and the primary outcomes pertain to the cluster level or the individual level.
Describe (1) the target population of clusters and individuals to which the study findings will be generalizable, and (2) the clusters to be randomized and the subjects to be enrolled in the trial.

RC-2: Justify the choice of cluster randomization.
Describe the benefits and disadvantages of cluster randomization versus individual-level randomization for the proposed research. Cluster randomization should be substantiated by a sound theoretical and conceptual framework that describes the hypothesized causal pathway (see **CI-1**). Cluster randomization generally is applicable in the following instances:

- An intervention is delivered at the cluster level.
- An intervention changes the physical or social environment.
- An intervention involves group processes.
- An intervention cannot be delivered without a serious risk of contamination.

Logistical considerations can also justify cluster randomization, for example, to reduce costs or to improve participation, adherence, or administrative feasibility.

RC-3: Power and sample size estimates must use appropriate methods to account for the dependence of observations within clusters and the degrees of freedom available at the cluster level.
The methods used to reflect dependence should be clearly described. Sources should be provided for the methods and for the data used to estimate the degree of dependence. Sensitivity analyses that incorporate different degrees of dependence must be reported. For simpler designs, the dependence in the data can be reflected in the intraclass correlation. Dependence can also be reflected in variance components. Other factors that affect the power calculation and should be described include the study design, the magnitude of the hypothesized intervention effect, the prespecified primary analysis, and the desired Type I error rate.

RC-4: Data analyses must account for the dependence of observations within clusters, regardless of its magnitude.
Data analyses must also reflect the degrees of freedom available at the cluster level. Investigators must propose appropriate methods for data analyses, with citations and sufficient detail to reproduce the analyses.

RC-5: Stratified randomization should be used when feasible.
Because cluster randomization trials often involve a limited number of groups or clusters, stratified randomization should be considered and is recommended when feasible. If not feasible, justification should be provided for the use of other methods. The recommended stratification factors are those that are expected to be strongly correlated with the outcome or with the delivery of the intervention, such as baseline value of the outcome variable, cluster size, and geographic area.

Only a limited number of confounders can be addressed through stratification. Other variables, particularly those that characterize the context, should be measured and assessed to document their potential influence on the outcome and understanding of heterogeneity of results.

13: STANDARDS FOR STUDIES OF COMPLEX INTERVENTIONS

SCI-1: Fully describe the intervention and comparator and define their core functions.

Describe the intervention and comparator under study and clearly define aspects related to core functions and forms. Core functions refer to the intended purpose(s) of the interventions. The form of the interventions includes the intended modes of delivery, providers involved, materials or tools required, dose, and frequency/intensity. The description should also explicitly indicate to whom the intervention is aimed (e.g., patient, provider, hospital, health system).

SCI-2: Specify the hypothesized causal pathways and their theoretical basis.

Clearly describe the hypothesized causal pathways by which the proposed complex intervention generates change (see **CI-1**). This description should depict how each intervention function generates the hypothesized effects on the prespecified patient outcome(s). Include in the causal model key contextual factors that may influence the impact of the intervention so that their hypothesized relationships are made explicit. Describe the theoretical and/or empirical bases underlying the proposed interventions and their hypothesized effects.

SCI-3: Specify how adaptations to the form of the intervention and comparator will be allowed and recorded.

Specify any allowable adaptations in form and describe how planned and unplanned adaptations will be managed, measured/documented, and reported over time. Any planned adaptations should have a clear rationale; be supported by theory, evidence, or experience; and maintain fidelity to the core functions of the intervention. Upon conclusion of the study, researchers should provide guidance on allowable adaptations or unproductive adaptations (i.e., adaptations that may reduce the effectiveness of an intervention).

SCI-4: Plan and describe a process evaluation.

Describe plans to conduct a process evaluation (i.e., to assess whether the intervention was implemented as planned and to test and refine the hypothesized causal pathways). Process evaluations should measure/document, analyze, and report the fidelity of the delivery of the intervention (i.e., planned and unplanned adaptations); the quantity or dose of the intervention actually delivered; whether the intended population(s) received the delivered intervention (i.e., reach); the mechanisms of action (e.g., mediators, intermediate outcomes); and important contextual factors (e.g., moderators), taking into account the levels at which the intervention is aimed (e.g., patient, provider, hospital).

Researchers should select a combination of methods appropriate to the process questions identified and describe the timing and sources of data collection. These plans should include appropriate quantitative, qualitative, and/or mixed methods that account for the intervention functions as defined by the causal pathway.

Describe the plans to integrate process and outcome data in advance of intervention delivery to determine whether and how outcomes and effects are influenced by implementation or contextual moderators. Explain how the results of the process evaluation will be used to draw inferences about both the effectiveness (i.e., patient outcomes) and the processes of care (i.e., process outcomes).

SCI-5: Select patient outcomes informed by the causal pathway.

Select valid and reliable patient outcome measures that are explicitly affected by the hypothesized causal pathway and the theoretical and/or empirical basis for the intervention. If the study does not measure a patient outcome, researchers must provide strong evidence that supports the linkage between the measured outcome and unmeasured patient outcome. The outcome measures should assess the intervention across a range of domains that sufficiently permit assessment of how the intervention affects patients. In determining the length of follow-up, assumptions about the rate and pattern of change expected in the outcome measures should be clear.

APPENDIX B: RESPONSE TO PUBLIC COMMENT

To promote transparency, meet legislative mandates, and increase the usefulness of the PCORI Methodology Standards, we use a formal process to solicit input from stakeholders. In preparing the 2018 update, we solicited public comments on a draft of the new standards for studies of complex interventions and for data integrity and rigorous analyses from October 30, 2017, to December 29, 2017.

We received comments from a broad spectrum of stakeholders, including health researchers, policy makers, and professional organizations. We thank the individuals and organizations that took time to provide the many thoughtful and meaningful suggestions.

After the comment period, the PCORI Methodology Committee and staff considered the submitted comments and made additional revisions to both the updated standards and the PCORI Methodology Report. The updated standards were adopted by PCORI's Board of Governors and posted at www.pcori.org in April 2018. The updated Methodology Report was posted in January 2019**.**

The table below displays all the public comments we received on the draft version of the updated standards. These comments have not been edited and are displayed in the table as they were submitted. The table also lists the stakeholder affiliation of the submitters (e.g., patient, health researcher) and our responses to each of the comments, including revisions to the standards or report.

Standard	Stakeholder Group	Comment (as submitted to PCORI)	Disposition
IR-7	Policy Maker	This seems like standard Data Safety and Monitoring Plan, often overseen by a DSMB.	Although a standard data safety and monitoring plan should ensure the integrity of the data collected, it does not generally cover the documentation and preservation of data for data-sharing purposes. PCORI believes that the additional requirements associated with a data management plan are necessary to ensure the integrity of the clinical research and support efforts related to open science.
	Health Researcher	Overall, the standards are quite useful and appear consistent with extant high quality data management approaches. However, for many implementation studies, it often is extremely difficult to conform to the high standards of a DMP as described. Therefore, the data collection and quality plan needs to be tested in the real setting in which it will be deployed. This should be done in advance so that problems can be foreseen and mitigated. Nowhere is this more important than in community based participatory research and other types of field research in low resource settings. Additionally, funders should acknowledge that adhering to these standards often requires substantial resources of time and money that should be allowed for in proposals.	Thank you for your comment. PCORI agrees that good data management practices are a requirement for all rigorously conducted research and that researchers should devote sufficient effort to ensuring adherence.

Standard	Stakeholder Group	Comment (as submitted to PCORI)	Disposition
SCI-1	Health Researcher	(1) AcademyHealth recommends that just as researchers should fully describe the intervention(s) and comparator(s) being studied, they should also describe the intervention implementation strategy, or at least the planned implementation approach. (2) When the comparator is the standard of care, this standard should be fully specified. (3) We also recommend that this standard include guidance for researchers to explain key contextual factors that may affect implementation outcomes, effectiveness, fidelity and variation across sites as well as a definition of the meaning of "levels". (4) Further we recommend that researchers describe the extent of freedom that exists for implementers to vary core functions and forms of the intervention, that is, how much variation in functions and forms is allowed within the study context (see SC-3 for more detail). In some instances (e.g. adaptive designs), researchers may be further refining the intervention mode of delivery through iterative phases and if so, this should be specified. (5) Minor edit on the last sentence in the standard: the examples in parentheses are just that – examples and this should be edited to be an e.g. not an i.e.	(1) **SCI-1** requires investigators to define aspects related to core functions and forms. The term "form" is used broadly and includes a description of how the intervention will be carried out to achieve its intended functions. (2) We agree that when the comparator is the standard of care, it should be fully specified as well as adequately justified. This point is addressed under the Standards for Formulating Research Questions (**RQ-5**). (3) We agree that it is important to outline (at the study design stage) how contextual factors might affect implementation as well as influence the outcomes. We address this in **SCI-2**, requiring investigators to illustrate how key contextual factors play a role in the causal pathway. **SCI-4** calls out that "levels" of an intervention should be described and accounted for in the process evaluation. The references provide further clarity on definitions and terminology. (4) We acknowledge the importance of specifying expected variation in intervention forms; however, investigators should ensure fidelity to core functions. **SCI-3** is intended to address this degree of standardization versus adaptation in form, and the report clarifies this point. (5) We edited this standard to state, "The description should also explicitly indicate to whom the intervention is aimed (e.g., patient, provider, hospital, health system)."

Standard	Stakeholder Group	Comment (as submitted to PCORI)	Disposition
SCI-1 (Continued)	Health Researcher	It is also important to describe how participants receive the components of the intervention. Are participants treated individually? In groups? By a common therapist or other change agent? Do they interact in groups or online? To the extent that participants interact with one another post randomization, whether face-to-face or electronically, we can expect some correlation in their data, and that would need to be addressed in the analysis. But if we don't track which participants are seen by which therapists, or receive components of the intervention in which groups, we cannot address it in the analysis.	**SCI-1** states that the description should explicitly indicate who the intervention is aimed at, which would include describing whether an intervention is directed toward a group of patients or individual patients. PCORI also has issued methodology standards on cluster research designs.
	Health Researcher	Perhaps add whether it is standardized or tailored, and how (although I do see that this is one of the characteristics above & may be what is referred to below as adaptations)	**SCI-3**, which requires investigators to specify adaptations, addresses the issue of standardizing versus tailoring an intervention.
	Health Researcher	Should the popn be under a different heading, and what about the setting?	In **SCI-1**, we require investigators to explicitly indicate to whom the intervention is aimed, which is intended to address the population. The Standards for Formulating Research Questions (i.e., **RQ-3** through **RQ-6**) and Standards Associated with Patient Centeredness (**PC-2**) address issues related to the study population and setting.
	Health Researcher	The description of the intervention should address key feasibility issues including likelihood of future implementation by stakeholders and effective use by patients.	We agree that addressing generalizability and implementation potential is important. **SCI-3** requires investigators to provide guidance based on study results of adaptations that may or may not be allowable. In addition, **PC-4** provides guidance on planning for the dissemination and implementation of study results.
	Policy Maker	Agree and very important to describe all intervention arms in detail. You don't indicate if this could be in supplemental materials or not, but most failures to describe interventions in detail are the result of inadequate space to do so in publications. Also, our terminology for describing intervention components is not standardized. I'd like to see some tip of the hat to efforts to standardize better these intervention components so we describe them similarly in publications.	The report that accompanies PCORI's Methodology Standards refers researchers to existing guidance and encourages researchers to use common terminology.
	Health Researcher	NO comment other than what is already listed in the Preamble section: specifically how different people with different roles, lived experiences, and training may be implementing the same actions in a complex intervention study.	We agree that there are often local adaptations for who is involved in implementing a complex intervention, which underscores the importance of specific training. **SCI-3** addresses adaptations in form, which includes how the intervention is delivered and who delivers it.

Standard	Stakeholder Group	Comment (as submitted to PCORI)	Disposition
SCI-1 (Continued)	Health Researcher	This definition is missing direction to describe the complexity itself, which is separate from the causal pathways in SCI-2. Simply stating the mode of delivery, providers, materials, dose, frequency, and target of the intervention not sufficient to help reviewers understand the complexity that should guide data collection and analyses. this description should further indicate the interaction/interdependency of any/all of the mentioned components on outcomes.	We agree that the interaction and interdependency of the functions and form of the intervention affect outcomes. **SCI-2** requires specification of the causal pathways, which is used broadly to imply the interdependence of various aspects of the intervention as well as contextual factors.
	Health Researcher	Nice distinction between form and function	Thank you for this comment.
	Health Researcher	In addition to describing the interventions, it is equally important to justify the choice for the intervention and control. Such choices should be based on several factors including (but not limited to): 1) Acceptability 2) Feasibility 3) Stringency 4) Uniformity 5) Relevance 6) Resemblance	We agree that the selection of intervention and comparator is critical. **RQ-5** addresses this issue, requiring investigators to describe why the comparators were selected, how they represent appropriate interventions in the context of the relevant causal model (**CI-1**), how they reduce the potential for biases, and how they allow direct comparisons.
SC-2	Health Researcher	Notes could suggest that it is not necessary, or even desirable, to base interventions on one theory. Simply, the rational for the presumed causal pathways should be described.	**SCI-2** requires the rationale for the causal pathway(s) to be supported by empirical evidence and/or theory. This standard does not imply that one theory would be sufficient to justify prespecified causal pathways
	Health Researcher	Consider specifying the direction of the hypothesized effects and rationale with appropriate support.	We agree that the direction of the relationships within the causal pathways are important and note this point in the report. Also note that the standard for causal inference methods (**CI-1**) requires that the causal model represent the potential mechanisms of effect and the conditions under which the hypotheses are to be tested
	Policy Maker	Excellent to require a causal pathway. That said, what often happens is that the investigator cites a theory or model, then fails to describe in detail how the various intervention component target causal mediators of that theory or model. Greater specificity is critical to this standard.	The PCORI Methodology Standards are intended to provide guidance but not be prescriptive. The report refers readers to more detailed guidance.

Standard	Stakeholder Group	Comment (as submitted to PCORI)	Disposition
SCI-2 (Continued)	Health Researcher	(1) AcademyHealth agrees that defining, in advance, the causal pathway and theoretical basis for change is necessary. To that end, the logic model should be supported by an established conceptual framework and appropriate citations provided. Logic models are linear and may fail to display the hypothesized interactions that are driven by context and complexity. Lack of a supporting conceptual model makes it difficult to see where the planned intervention fits in the larger context, as well as to visualize interactions. (2) Accounting and planning for context is critical as context inevitably has an impact on the dose of an intervention that is actually received, which could be seen as a mediator of outcomes. Thus, the context influences the actual intervention (forced modification of the intervention due to the context), the dose of the planned intervention that actually is received, the success of the planned activities in achieving planned outputs, change in process measures/behaviors, and outcomes. These dimensions should be mapped to the components of the logic model for clarity and not be limited to "prespecified patient outcomes(s)". (3) In addition, to recognize the potential to ensure maximum learning from the implementation of the intervention, we recommend requiring both the documentation of the causal pathway in advance (prior to seeing the data) as well as any modifications made to the model after data analysis and the rationale for such changes. Retrospective analysis, and publication, of such discovered pathways and the rationale for the changes in the logic model will further contribute to the body of knowledge. (4) We further note a theory explaining how and why the proposed interventions will affect the outcomes is insufficient. Researchers should provide an explicit quantitative prediction of the attributable effect, along with the expected precision of this estimate (in the form of degree of belief, prior probability, or confidence intervals). Too often failure to specify a target outcome is due to lack of good evidence regarding the likely attributable effect, failure to consider the low reliability of health systems in implementing core changes, and/or adequate consideration of bias and confounding. The hypothetical causal pathway and the implied attributable effect should be weighted in the light of the Bradford Hill or other epidemiological criteria/standards. (5) Minor edit to the third sentence: it would be unrealistic to ask for "any" contextual factors. This should refer to "key" contextual factors.	(1) The PCORI Methodology Standards are intended to provide guidance but not be prescriptive. **SCI-2** requires investigators to depict their causal pathway. The report refers readers to more detailed guidance. (2) **SCI-2** recognizes the complex, and often indirect, relationship between contextual factors and patient outcomes. The standard requires investigators to include key contextual factors in the causal model but does not imply a direct relationship between contextual factors and patient outcomes. (3) **SCI-4** requires investigators to use the results of the study and process evaluation to inform the hypothesized causal model. (4) We agree about the importance of providing both empirical and theoretical support for a hypothesized causal pathway. **SCI-2** requires researchers to support their rationale for the causal pathway(s) with empirical evidence and/or theory. (5) The standard has been revised to incorporate this suggestion.

Standard	Stakeholder Group	Comment (as submitted to PCORI)	Disposition
SCI-3	Health Researcher	(1) Recognizing that an intervention will not be delivered 100 percent of the time to 100 percent of the patients or community members—and often for good reasons—AcademyHealth recommends that this standard not only detail specification of adaptations, but also the documentation of unplanned, observed adaptations that were not pre-specified, and the rationale, setting, and frequency of those adaptations. (2) The need to describe clearly the rationale for any adaptation goes beyond the desirability of comprehensiveness or completeness. In cases where an adaptation to an intervention may be chosen because the chooser knows or suspects that this particular adaptation will work better than any other in the specific setting, outcomes of the particular adaptation may be uniquely good in the setting in question, but not generalizable to other settings. Allowing for endogenous adaptations may itself be a characteristic of an intervention, but this needs to be appreciated and documented. Qualitative data is especially helpful in understanding whether an adaptation was chosen because it was known or suspected to be especially effective in a specific setting.	(1) We agree that the rationale, setting, and frequency of both planned and unplanned (observed) adaptations should be documented. **SCI-3** has been revised to state that researchers should specify how both planned and unplanned adaptations will be managed, measured/documented, and reported over time. Researchers may consult the citations in the report for further guidance. (2) We agree that outlining the rationale for any adaptation is critical and that some adaptations may not be generalizable. **SCI-3** asks researchers to provide guidance on adaptations upon conclusion of the study, which may include the site-specific adaptations in relation to what is and what isn't generalizable to all settings. **SCI-4** addresses the importance of collecting qualitative data to help better understand implementation.
	Policy Maker	This is also an excellent standard that encourages planned adaptations.	Thank you for this comment.
	Health Researcher	I think another issue here is the intended degree of pragmatism of the trial. Perhaps, using the PRECIS model would be helpful.	We agree that the extent to which planned adaptations are allowed is related to how tightly an intervention is defined. Regardless of the amount of flexibility or pragmatism afforded to an intervention, adaptations should be explicitly considered and specified ahead of time to ensure a well-defined intervention and adequate data collection approaches to track implementation, fidelity, and unplanned adaptations. The report outlines these points.
	Health Researcher	"Agree! My only thought here is that efficient and easily understandable data systems must be built that can handle this type of longitudinal information collection strategy. And folks with biostatistical know-how need to be on these teams such that such changes can be handled in the analysis. Research teams have to lead these types of efforts, clinical staff cannot be relied upon to try to think of and document staff changes, role changes and other changes in the context of research.	We agree that efficient studies minimize the burden on clinical staff for data collection. **IR-2** requires investigators to assess data source adequacy.

Standard	Stakeholder Group	Comment (as submitted to PCORI)	Disposition
SCI-4	Training Institution	There is an opportunity here to recommend the use of statistical process control charts for analysis of complex interventions. SCI-4 rightly describes matching the analysis to the questions, but with complex interventions the emerge and change over time, SPC is the best method to monitor the effects on the system. I recognize the likely hesitancy to recommend any specific type of analysis, but perhaps a statement such as "methods used to draw inferences from the data on efficacy and understand the variation of outcomes over time"	We agree that complex interventions evolve and therefore outcomes may vary over time. **SCI-4** requires investigators to "describe the timing and sources of data collection . . . determine whether and how outcomes and effects are influenced by implementation or contextual moderators" and the text of the report provides additional guidance about approaches to capture such changes.
	Health Researchers	Other methods such as Qualitative Comparative Analysis (QCA) and Configurational Comparative Methods (CCM) are valid and more robust than quantitative methods for determining causal pathways in complex interventions and systems. Because they are complex, standard quantitative methods alone are not sufficient for understanding and evaluating complex interventions. The third paragraph above should state instead in the second sentence that plans should include appropriate quantitative, qualitative, and mixed methods analyses. In complex interventions, qualitative and mixed methods analyses are NOT supplemental to quantitative methods - they are critical to understanding and evaluation in these interventions. Quantitative analyses alone are insufficient for evaluating and understanding complex interventions, yet the third paragraph as currently written is heavily weighted that direction.	The standard has been revised to incorporate this suggestion and emphasize the importance of qualitative approaches.
	Health Researcher	It should be recognised that formal measurement of mediating, intermediate, outcomes is not always possible. Measures are not always available, and if they are available participant burden can often be far too high. I have experience of working in very deprived areas in the UK and in low literacy settings in Low and Middle Income Countries. In these settings it is imperative to minimise respondent burden. So we need to recognise that it is not always practically possible to gather sufficient data to full test the causal pathways.	We agree that measurement of outcomes should be tailored to and appropriate for an intervention's setting, and that participant burden should be considered. The PCORI Methodology Standards discuss the selection of outcomes (**RQ-6**) and patient centeredness (**PC-1** through **PC-4**) in detail. The Methodology Report also addresses this issue.
	Health Researcher	I think gender/sex interactions with interventions should be explored whenever relevant and possible. Also, the guidance can be more specific about how to approach subgroup analyses to minimize false positives and false negatives.	The Standards for Heterogeneity of Treatment Effects and accompanying text provide detailed guidance for prespecifying subgroup hypotheses and approaches for analyzing data about subgroup effects.

Standard	Stakeholder Group	Comment (as submitted to PCORI)	Disposition
SCI-4 (Continued)	Health Researcher	Process evaluations should be required. Often this evaluation involves use of mixed methods to inform why or why not the complex intervention succeeded (or not) by examining causal pathways often in the form of logic model that addresses key steps in the process. Given that complex interventions often fail, such data are vital to informing next research steps and enhance the value of "negative studies." Process evaluation is analogous to but not synonymous with testing of basic hypothesized mechanisms. In the case of process evaluation of complex interventions, the goal is to inform to determine whether the steps in the process occurred as anticipated in this particular study. The latter (testing of hypothesized basic mechanisms) is designed to generate generalizable knowledge regarding fundamental physiological or behavioral mechanisms.	**SCI-4** requires investigators to plan and describe a process evaluation, and we revised the text to reflect this. The standard has also been revised to emphasize the importance of the use of qualitative and mixed methods. We agree that process evaluation is not synonymous with testing causal mechanisms; however, results of the process evaluation can be used to inform uncertainties in the causal pathways.

Standard	Stakeholder Group	Comment (as submitted to PCORI)	Disposition
SCI-4 (Continued)	Policy Maker	This looks like what should be routine protocol specification. This standard seems broad enough that most investigators will be able to say they already do this. Not sure if there is something specific here that the standard is trying to achieve that is not already commonly achieved. Greater specificity seems needed for this standard	We revised the title of the standard to clarify its focus on process evaluation. Greater specificity may be considered in future revisions to the standards.
		Agree, but I also think it behooves the funding agencies to share models that they feel are particularly valuable at this point in our evolution of patient centered research. At least some examples of thoughts on constructing conceptual models, but indeed with a keen eye on driving research teams to consider if their choices truly fit within a larger conceptual framework vs. just coming up with things that may be interesting, but ultimately unconnected and over burdensome to all from a volume perspective. Research teams should explain their plans for how they will message about their studies to clinical staff (be sense makers) and how they will continue to be available to clinical staff to keep them informed and on task.	Thank you for this comment. We will consider your suggestion in future revisions to the standards.
	Health Researcher	the process evaluation will not always be able (nor will it be appropriate to measure all of these things - this should be made clear).	We agree that data collection approaches need to consider feasibility and respondent burden. The PCORI Methodology Standards discuss the selection of outcomes (**RQ-6**) and patient centeredness (**PC-1** through **PC-4**) in detail. The report addresses this issue as well.
	Health Researcher	Complex interventions are often multi-level, and variables are often measured at multiple levels. It is common to evaluate multi-level interventions with group- or cluster-randomized designs, stepped wedge designs, or with individually randomized group-treatment or partially clustered designs. Such designs pose special sample size and analytic issues, and these would need to be addressed in the application. The important point is to anticipate whether observations will be correlated, and to address that correlation when the study is being planned and analyzed. A useful resource is available from NIH at https://researchmethodsresources.nih.gov.	We agree with these points. PCORI has released Standards on Research Designs Using Clusters that include guidance on when cluster designs are appropriate.
	Health Researcher	The phrase “nature of the functions defined by the causal pathways” was not clear to me.	The standard has been revised to read “intervention functions as defined by the causal pathway.”

Standard	Stakeholder Group	Comment (as submitted to PCORI)	Disposition
SCI-4 (Continued)	Health Researcher	(1) The wording of this standard raises a number of concerns. First, effectiveness can be measured by both process and outcomes. The term "process outcome" is confusing, however we recognize that in some instances, especially when true outcomes are hard to obtain, intermediate outcomes are used, and these can in fact be processes. We suggest editing the first sentence to: "draw inferences about the impact of the intervention on processes of care and patient outcomes". (2) Second, we strongly support the use of valid and reliable patient outcome measures but only when they are appropriate to the patient, population, intervention and context. The standard should not encourage use of measures for a different patient population/setting that the researchers think are inappropriate for their study. This statement also should discuss the balance between process and outcome measures. When outcomes are rare, hard to capture, or in the distant future, an explicit case must be made for why process measures are reasonable proxies. (3) Third, the statement also appears to favor quantitative methods over qualitative ones. AcademyHealth suggests that for complex interventions qualitative and mixed methods should be more strongly suggested as quantitative methods alone are likely to be insufficient. In fact, some questions may only be analyzed with rigorous qualitative methods. (4) Fourth, specifying contextual factors at all levels of the targets of the intervention is certainly the ideal; however, depending upon the scope and budget of the project, it may be reasonable to target measures to selected levels or selected aspects of the process. (5) Fifth, Researchers should describe in detail how the subject/settings for study were identified and how intervention status was assigned. In many health services research studies of complex interventions the subjects may be a convenience sample and intervention assignment may be based on voluntary participation. In others, the data will be observational with or without a true natural experiment. In either instance, researchers should identify and describe potential sources of bias and, if possible, determine the likely direction of the bias. Researchers should also describe the methods used to minimize bias and to quantify its likely magnitude and direction. Additional points include: (6) It would be helpful to define the expected duration of the intervention as well as the expected timeline for effects to appear for various outcomes and processes. This may well differ for different settings, patients, and populations and researchers should describe how this will be determined. Effects may appear at different times for different outcomes and processes. In complex interventions, it may take a considerable amount of time for the program to become fully effective—researchers should specify that for each outcome or group of outcomes and the basis for their assumptions. (7) The definition of data collection tools and sources should be documented and additional implementation outcomes (acceptability, reach) and strategy outcomes (speed, quality, reach) should be considered. An assessment of the strength of evidence for the anticipated impact on outcomes is also necessary. (8) It would be wise to consider collecting data on the potential costs and budget impact of the intervention and its implementation, including ongoing maintenance and opportunity costs of the intervention. Expending significant effort to develop and test a complex intervention that has little to no likelihood of adoption because of its cost is not ideal.	The standard has been revised to focus on process evaluation. (1) We agree that, in some cases, effectiveness may be measured by both process and outcomes. **RQ-6** requires that studies select outcomes that are patient centered. (2) We agree that outcome measures should be appropriate for the patient population, intervention, and context. We added **SCI-5** to address this issue. Also, **RQ-6** discusses the selection of outcomes that people representing the population of interest notice and care about (i.e., patient outcomes). (3) The standard has been revised to read "... include appropriate quantitative, qualitative, and/or mixed methods ..." (4) We agree that it may be reasonable to target measures to selected contextual factors. Therefore, we revised this statement to read, "important contextual factors (e.g., moderators) taking into account the levels at which the intervention is aimed." (5) The Standards for Formulating Research Questions (**RQ-1** through **RQ-6**) and Standards Associated with Patient Centeredness (**PC-1** through **PC-4**) direct researchers to describe why specific patient populations and settings were chosen and to describe how participants are identified, selected, recruited, and enrolled. **CI-1**, **CI-2**, and **CI-4** provide guidance on identifying, describing, minimizing, and quantifying bias. Additional points: (6) Standard **SCI-1** instructs researchers to describe the frequency/intensity of the intervention, and the Standards for Causal Inference Methods (**CI-3**) ask researchers to describe the timing of the outcome assessment relative to the initiation and duration of the exposure. (7) The Standards for Data Integrity and Rigorous Analyses provide guidance for defining data collection tools and sources (**IR-1**to **IR-4**). (8) We agree that it is important to document barriers of any variety to the adoption of interventions; however, discussing the collection of cost data is beyond the scope of these standards.

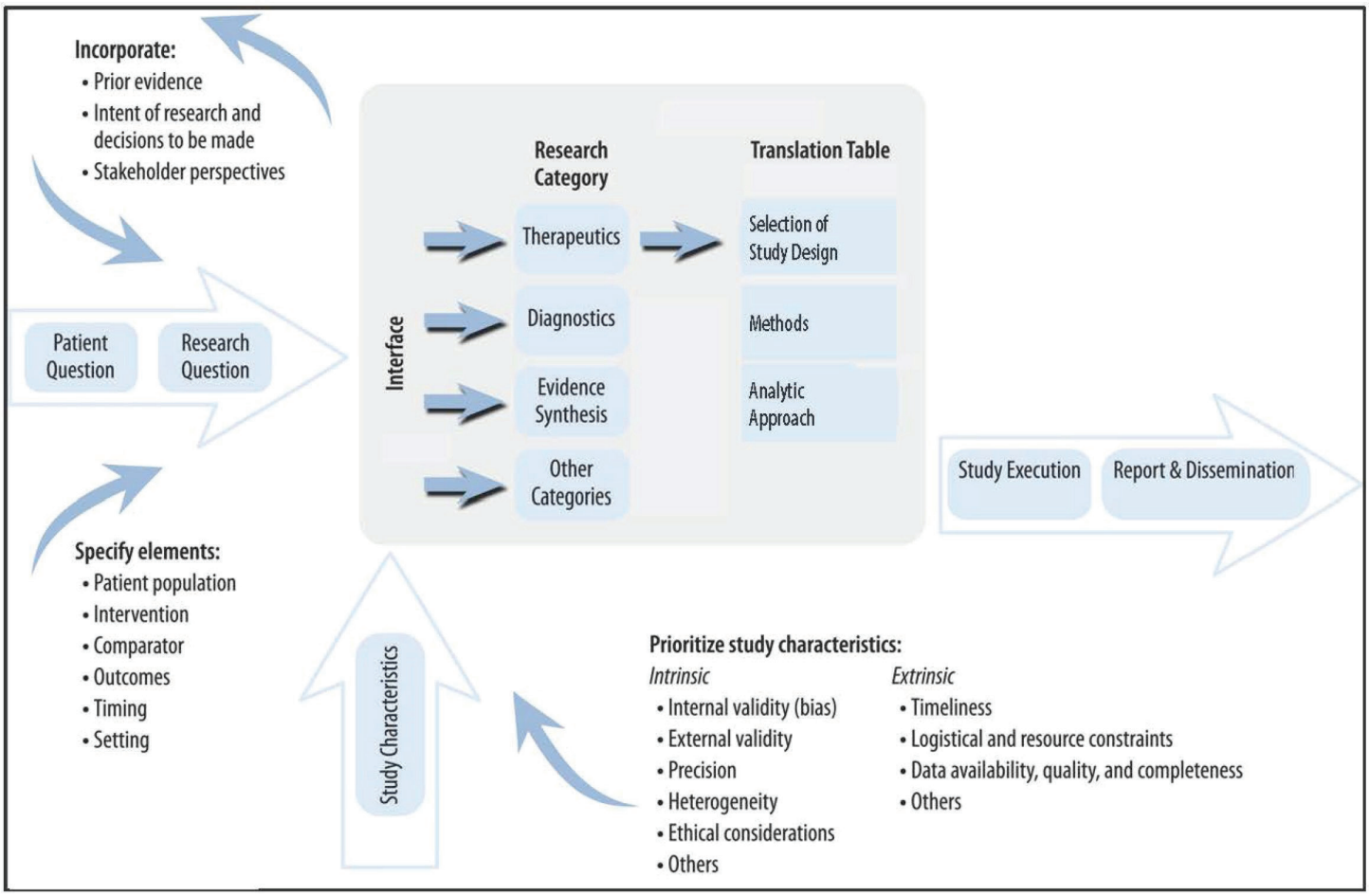
Standard	Stakeholder Group	Comment (as submitted to PCORI)	Disposition
General Comments	Health Researcher	AcademyHealth recommends that PCORI be more specific about what "levels" are being referred to in "multiple entities or levels targeted by the intervention". Is this referring to the six levels of community, patient, provider, microsystem, mesosystem, and macrosystem. A clear definition should be provided.	The "preamble" is a part of the methodology report, which provides the rationale underlying the standards. Researchers may consult the citations throughout the report for further clarity.
	Health Researcher	not just healthcare staff	We have edited the preamble to read "individuals (e.g., healthcare staff, providers, patients, caregivers)."
	Policy Maker	Very useful to identify this issue. Multiple active ingredients has been dealt with by behavioral interventionists for decades.	Thank you for this comment.

Standard	Stakeholder Group	Comment (as submitted to PCORI)	Disposition
General Comments (Continued)	Health Researcher	(1) It is also important to point out that in different healthcare settings, the people actually fulfilling the roles required to implement and intervention may be different from setting to setting. Thus, a medical assistant may be performing actions in one setting where in another setting that same work is being performed by an registered nurse, a physician, or even a community stakeholder. This is why specific training is so important, but training that does not expect that all those involved are necessarily starting from the same place. As well activities performed by one person in one setting may also be done by more than one person in another. (2) It would be good to consider how to describe and engage in complex interventions such that some activities do not need to be performed with fidelity while others do. Thus, have adaptable and nonadaptable components. Investigators could set a priori which interventions or actions need to be carried out with fidelity vs. which components could be more adaptable. For instance in a practice or health system level intervention that is dependent on continually identifying the cohort of subjects to recruit, the practice or health system needs to commit to having dedicated staff members serve roles in pulling patient cohorts in a standard and consistent query throughout the trial. As well to mitigate measurement bias, investigators could describe why fidelity to certain measurements, like accurate assessment of blood pressure of participants in a hypertension trial, is critical to the scientific integrity of the trial. (3) Also there likely needs to be some thought put into the issue of how Vanguard site experiences are used to inform trials. Are there standards for this? In many pragmatic trials, the teams start with Vanguard sites to work out many details, but in some cases the Vanguard site continues to enroll subjects, implement protocols while the intervention sites (nonvanguard) are activated. How are lessons from the Vanguard woven into protocols/decisions made by teams? How is this issue of temporality handled when there is overlap between Vanguard and intervention phase sites? Is there some kind of in analytical approach needed in such cases?	Thank you for this comment. (1) We agree that there are often local adaptations for who is involved in implementing a complex intervention, which underscores the importance of specific training. **SCI-3** addresses adaptations in form, which includes how the intervention is delivered and who delivers it. (2) We agree that complex interventions have adaptable and nonadaptable components. **SCI-3** addresses this point by requiring investigators to specify any allowable adaptations in form. Investigators should preserve fidelity to functions. (3) The standards do not explicitly address the role of vanguard sites to inform trials. In the context of studies of complex interventions, vanguard sites usually would inform how to implement permissible adaptations in form across study sites. Specifying allowable adaptations is addressed in **SCI-3**. Vanguard sites may also be a component of process evaluations, which are addressed in **SCI-4**.

Standard	Stakeholder Group	Comment (as submitted to PCORI)	Disposition
General Comments (Continued)	Health Researcher	AcademyHealth believes the pre-definition and documentation of planned analyses, data sources, and data collection tools as outlined in these standards is highly desirable, and will support the overall integrity of the research. We also acknowledge that doing so requires a significant investment of time and budget. Successful adherence to these standards may be difficult or unrealistic for a study with a limited budget, and considerations for this work should be made in the grant process. To the degree possible, we also recommend simplification of the language and word choice in each standard to ensure both specificity and understanding across settings and disciplines. Finally, it is not clear whether this guidance has been cross-walked with the published guidance in SQUIRE and STaRI and a crosswalk would be a helpful table, and if there are gaps in these guidelines or in SQUIRE and STaRI, they should be addressed.	We will take this into consideration. We revised the standards to address lack of clarity and simplification identified in the public comments. Extensive guidance was reviewed and cross-walked in the process of developing the standards. We reference key guidance applicable to PCOR in the PCORI Methodology Report.
	Health Researcher	You still are missing the boat by not including qualitative research standards in your methodology report. Please strongly consider this. I am an expert in qualitative research, and would be happy to work with you as a consultant to help develop rigorous, patient-centered standards.	PCORI has undertaken an effort to develop standards on qualitative and mixed methods. These will be added in future revisions of the standards.
	Health Researcher	Nonadherence to treatment can be a particularly serious concern in complex interventions. How will this be addressed, e.g., ITT analysis or per-protocol analysis, or something else? This needs to be clearly stated and justified. If this is already addressed in a different standard, one could point to that standard. Another issue is that blinding is typically impossible in complex interventions. What would be the impact of lack of blinding (e.g., placebo effect?)	We agree that issues of nonadherence and masking are important to complex interventions. **IR-1** requires investigators to specify plans for quantitative data analysis that correspond to major aims. The PCORI Methodology Standards are intended to be minimal and not prescriptive. **IR-6** states that when masking is not feasible, the impact of lack of masking on the results should be discussed.
	Health Researcher	The standards are comprehensive - but the data collection and analysis standards might prove unachievable as written	As mentioned under **SCI-4**, **RQ-6** emphasizes the importance of ensuring outcomes are selected with patient centeredness (and, thus, participant burden) in mind.
	Policy Maker	These are excellent standards that are great to see PCORI considering.	Thank you for this comment.
	Health Researcher	This newly proposed content is logical, but it would be great for PCORI to publish a template for investigators to see that demonstrates how to craft this language. The challenge is often that the funding agency wants more description of actions that research team will take, but then still wants to limit the number of pages for proposals. So please be careful about how much more you are asking for if you are keeping the page limits the same.	We will take this into consideration.

Standard	Stakeholder Group	Comment (as submitted to PCORI)	Disposition
General Comments (Continued)	Health Researcher	Thank you for providing updated methodology standards specific to complex interventions and recognizing the important differences	Thank you for this comment.
	Health Researcher	These are important improvements in the standards for PCORI methods. They still stop one step short of enabling vigorous learning organizations. PCORI methodology committee should undertake a serious study of Shewhart statistics and its approach to claims of changes worth understanding, which are grounded in quite reputable statistics. In addition, when getting into the effects of context, PCORI methodologists should understand and be willing to use the Context-Mechanism-Outcome structure now widely used in Britain and Europe and initially spelled out by Pawson and Tilley. CMO combinations that are highly context-dependent will challenge our wisdom as to generalizability, but Shewhart statistics can provide solid guidance as to whether the effects upon outcomes are important to understand.	We will take this into consideration in future revisions to the methodology standards.
	Health Researcher	Useful and rigorous.	Thank you for this comment.

APPENDIX C: TRANSLATION FRAMEWORK



APPENDIX D: REFERENCES

Agency for Healthcare Research and Quality. Outcome measures framework: literature review findings and implications. Registry of patient registries. Retrieved from https://www.effectivehealthcare.ahrq.gov/ehc/products/311/2310/registry-of-patient-registries-OMF-report-160912.pdf. Published 2016.

American Institutes for Research. PROMIS: dynamic tools to measure health outcomes from the patient perspective. Retrieved from http://www.nihpromis.com/. Published 2016.

Bauman LJ, Stein, RE, Ireys HT. Reinventing fidelity: the transfer of social technology among settings. *Am J Community Psychol.* 1991;19(4):619-639.

Benchimol EI, Smeeth L, Guttmann A, et al. The reporting of studies conducted using observational routinely-collected health data (RECORD) statement. *PLoS Med*. 2015;12(10):e1001885.

Berry SM, Carlin BP, Lee JJ, Muller P. *Bayesian Adaptive Methods for Clinical Trials*. Boca Raton, FL: CRC Press; 2010.

Blackwood B, Alderdice F, Burns K, Cardwell C, Lavery G, O'Halloran P. Use of weaning protocols for reducing duration of mechanical ventilation in critically ill adult patients: Cochrane systematic review and meta-analysis. *BMJ*. 2011;342:c7237.

Bonell C, Fletcher A, Morton M, Lorenc T, Moore L. Realist randomised controlled trials: a new approach to evaluating complex public health interventions. *Soc Sci Med*. 2012;75(12):2299-2306.

Bossuyt PM, McCaffery K. Additional patient outcomes and pathways in evaluations of testing. *Med Decis Making*. 2009; 9(5): E30-38.

Boutron I, Altman DG, Moher D, Schulz KF, Ravaud P. CONSORT statement for randomized trials of nonpharmacologic treatments: a 2017 update and a CONSORT extension for nonpharmacologic trial abstracts. *Ann Intern Med*. 2017;167(1):40-47.

Brennan PF, Stead WW. Assessing data quality: from concordance, through correctness and completeness, to valid manipulatable representations. *J Am Med Inform Assoc*. 2000;7(1):106-107.

Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey Smith G. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. *Health Technol Assess.* 2001;5(33):1-56.

Byng R, Norman I, Redfern S, Jones R. Exposing the key functions of a complex intervention for shared care in mental health: case study of a process evaluation. *BMC Health Serv Res*. 2008;8(1):274.

Campbell MK, Piaggio G, Elbourne DR, Altman DG. Consort 2010 statement: extension to cluster randomised trials. *BMJ*. 2012; 345:e5661.

Carlson JJ, Thariani R, Roth J, et al. Value-of-information analysis within a stakeholder-driven research prioritization process in a US setting: an application in cancer genomics. *Med Decis Making*. 2013;33(4):463-471.

Carman KL, Dardess P, Maurer M, et al. Patient and family engagement: a framework for understanding the elements and developing interventions and policies. *Health Aff*. 2013;32(2), 223-231.

CAST-II (Cardiac Arrhythmia Suppression Trial-II) Investigators. Effect of antiarrhythmic agent moricizine on survival after myocardial infarction. *New Engl J Med*. 1992;327(4):227-233.

Chalmers I, Bracken MB, Djulbegovic B, et al. How to increase value and reduce waste when research priorities are set. *Lancet*. 2014;383(9912):156-165.

Chalmers I, Glasziou P. Avoidable waste in the production and reporting of research evidence. *Lancet*. 2009;374(9683):86-89.

Chan A, Song F, Vickers A, et al. Increasing value and reducing waste: addressing inaccessible research. *Lancet*. 2014;383(9913):257-266.

Claxton KP, Sculpher MJ. Using value of information analysis to prioritise health research: some lessons from recent UK experience. *Pharmacoeconomics*. 2006;24(11):1055-1068.

Coly A, Parry G. Evaluating complex health interventions: a guide to rigorous research designs. AcademyHealth; June 2017.

Craig P, Dieppe P, Macintyre S, Michie S, Nazareth I, Petticrew M. Developing and evaluating complex interventions: the new Medical Research Council guidance. *Int J Nurs Stud*. 2013;50(5):587-592.

Crandall W, Kappelman MD, Colletti RB, et al. ImproveCareNow: the development of a pediatric inflammatory bowel disease improvement network. *Inflamm Bowel Dis*. 2011:17(1):450-457.

Dahabreh IJ, Hayward R, Kent DM. Using group data to treat individuals: understanding heterogeneous treatment effects in the age of precision medicine and patient-centered evidence. *Int J Epidemiol*. 2016. doi:10.1093/ije/dyw125.

Detre K, Peduzzi P, Murphy M, et al. Effect of bypass surgery on survival in patients in low- and high-risk subgroups delineated by the use of simple clinical variables. *Circulation*. 1981;63(6):1329-1338.

Diaz-Ordaz K, Froud R, Sheehan B, Eldridge S. A systematic review of cluster randomised trials in residential facilities for older people suggests how to improve quality. *BMC Med Res Methodol*. 2013;13(1):127.

Domecq JP, Prutsky G, Elraiyah T, et al. Patient engagement in research: a systematic review. *BMC Health Serv Res*. 2014;14(1):89.

Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. New York, NY: Oxford University Press; 2010.

Dudley L, Gamble C, Preston J, Buck D, Hanley B, Williamson P. What difference does patient and public involvement make and what are its pathways to impact? Qualitative study of patients and researchers from a cohort of randomised clinical trials. *PLoS One*. 2015;10(6):e0128817.

Echt DS, Liebson PR, Mitchell LB, et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo—the cardiac arrhythmia suppression trial. *New Engl J Med*. 1991;324(12),781-788.

Emsley R, Dunn G, White IR. Mediation and moderation of treatment effects in randomised controlled trials of complex interventions. *Stat Methods Med Res*. 2010;19(3):237-270.

Esmail L, Moore E, Rein A. Evaluating patient and stakeholder engagement in research: moving from theory to practice. *J Comp Eff Res*. 2015;4(2):133-145.

Evans I, Thornton H, Chalmers I, Glasziou P. *Testing Treatments: Better Research for Better Healthcare*. 2nd ed. London, UK: Pinter & Martin; 2011.

Ferrante di Ruffano L, Hyde CJ, McCaffery KJ, Bossuyt PM, Deeks JJ. Assessing the value of diagnostic tests: a framework for designing and evaluating trials. *BMJ*. 2012;344:e686.

Fiore LD, Brophy M, Ferguson RE, et al. A point-of-care clinical trial comparing insulin administered using a sliding scale versus a weight-based regimen. *Clin Trials*. 2011;8(2):183-195.

Fleurence RL, Forsythe LP, Lauer M, et al. Engaging patients and stakeholders in research proposal review: the Patient-Centered Outcomes Research Institute. *Ann Intern Med*. 2014;161(2):122-130.

Forsythe LP, Ellis LE, Edmundson L, et al. Patient and stakeholder engagement in the PCORI pilot projects: description and lessons learned. *J Gen Intern Med*. 2016;31(1),13-21.

Frank L, Basch E, Selby JV. The PCORI perspective on patient-centered outcomes research. *JAMA*. 2014;312(15),1513-1514.

Frank L, Forsythe L, Ellis L, et al. Conceptual and practical foundations of patient engagement in research at the Patient-Centered Outcomes Research Institute. *Qual Life Res*. 2015;24(5):1033-1041.

Gagnon M, Desmartis M, Lepage-Savary D, et al. Introducing patients' and the public's perspectives to health technology assessment: a systematic review of international experiences. *Int J Technol Assess Health Care*. 2011;27(01):31-42.

Glasziou P, Altman DG, Bossuyt P, et al. Reducing waste from incomplete or unusable reports of biomedical research. *Lancet*. 2014;383(9913):267-276.

Goeschel CA, Pronovost PJ. Harnessing the potential of health care collaboratives: lessons from the Keystone ICU Project. In: Henriksen K, Battles JB, Keyes MA, et al., eds. *Culture and Redesign*. Rockville, MD: Agency for Healthcare Research and Quality; 2008. *Advances in Patient Safety: New Directions and Alternative Approaches*; vol. 2.

Goldfine AB, Kaul S, Hiatt WR. Fibrates in the treatment of dyslipidemias—time for a reassessment. *New Engl J Med*. 2011;365(6):481-484.

Goodman SN, Schneeweiss S, Baiocchi M. (2017). Using design thinking to differentiate useful from misleading evidence in observational research. *JAMA*. 2017;317(7):705-707.

Greenhalgh T, Robert G, Macfarlane F, Bate P, Kyriakidou O. Diffusion of innovations in service organizations: systematic review and recommendations. *Milbank Q*. 2004;82(4):581-629.

Guise JM, Chang C, Butler M, Viswanathan M, Tugwell P. AHRQ series on complex intervention systematic reviews—paper 1: an introduction to a series of articles that provide guidance and tools for reviews of complex interventions. *J Clin Epidemiol*. 2017;90:6-10.

Hawe P. Lessons from complex interventions to improve health. *Annu Rev Public Health*. 2015;36:307-323.

Hawe P, Shiell A, Riley T. Complex interventions: how "out of control" can a randomised controlled trial be? *BMJ*. 2004;328(7455): 1561.

Helfand M, Tunis S, Whitlock EP, Pauker SG, Basu A, Chilingerian J. A CTSA agenda to advance methods for comparative effectiveness research. *Clin Transl Sci*. 2011;4(3):188-198.

Hernán MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. *Epidemiology*. 2008;19:766-779.

Hess EP, Knoedler MA, Shah ND, et al. The chest pain choice decision aid: a randomized trial. *Circ Cardiovasc Qual Outcomes*. 2012;5(3):251-259.

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 Intern Med*. 2011;154(8):509-515.

Hoffmann TC, Glasziou PP, Boutron I, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*. 2014;348:1687.

Institute of Medicine. *Knowing What Works in Health Care: A Roadmap for the Nation*. Washington, DC: The National Academies Press; 2008.

Institute of Medicine. *Initial National Priorities for Comparative Effectiveness Research*. Washington, DC: The National Academies Press; 2009.

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

Institute of Medicine. *Ethical and Scientific Issues in Studying the Safety of Approved Drugs*. Washington, DC: The National Academies Press; 2012.

Institute of Medicine. *Observational Studies in a Learning Health System: Workshop Summary*. Washington, DC: The National Academies Press; 2013.

Ioannidis JP, Greenland S, Hlatky MA, et al. Increasing value and reducing waste in research design, conduct, and analysis. *Lancet*. 2014;383(9912):166-175.

ISIS-1 (First International Study of Infarct Survival) Collaborative Group. Randomized trial of intravenous atenolol among 16027 cases of suspected acute myocardial infarction: ISIS-1. *Lancet*. 1986;2(8498):57-66.

Kahn M, Eliason B, Bathurst J. Quantifying clinical data quality using relative gold standards. *Proceedings of the AMIA Annual Symposium*. 2010;356-360.

Kahn MG, Batson D, Schilling LM. Data model considerations for clinical effectiveness researchers. *Med Care*. 2012;50:S60–S67.

Kahn MG, Brown JS, Chun AT, et al. Transparent reporting of data quality in distributed data networks. *eGEMs*. 2015;3(1):7.

Kahn MG, Callahan TJ, Barnard J, et al. A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *eGEMs*. 2016;4(1):18.

Kane H, Lewis MA, Williams PA, Kahwati LC. Using qualitative comparative analysis to understand and quantify translation and implementation. *Transl Behav Med*. 2014;4(2):201-208.

Kent DM, Rothwell PM, Ioannidis JP, Altman DG, Hayward RA. Assessing and reporting heterogeneity in treatment effects in clinical trials: a proposal. *Trials*. 2010;11(1):85.

Kent DM, Steyerberg EW, van Klaveren D. (2018). Personalized evidence-based medicine: predictive approaches to heterogeneous treatment effects. *BMJ* 364:k4245.

Kim, L.G., P Scott, R.A., Ashton, H.A., Thompson, SG. A sustained mortality benefit from screening for abdominal aortic aneurysm. *Ann Intern Med*. 2007;146(10):699-706.

Koepsell TD. Epidemiologic issues in the design of community intervention trials. In: Brownson RC, Petitti DB, eds. *Applied Epidemiology: Theory to Practice*. New York, NY: Oxford University Press; 1998:177-211.

Kotchen TA, Spellecy R. Peer review: a research priority. White paper. Retrieved from http://www.pcori.org/ assets/Peer-Review-A-Research-Priority.pdf. Published 2012.

Lagakos SW. The challenge of subgroup analyses—reporting without distorting. *New Engl J Med*. 2006;354(16):1667–1669.

Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. *Implement Sci*. 2010;5(1):69.

Li T, Hutfless S, Scharfstein DO, et al. Standards should be applied in the prevention and handling of missing data for patient-centered outcomes research: a systematic review and expert consensus. *J Clin Epidemiol*. 2014;67(1):15-32.

Lipitz-Snyderman A, Steinwachs D, Needham DM, Colantuoni E, Morlock LL, Pronovost PJ. Impact of a statewide intensive-care unit quality-improvement initiative on hospital mortality and length of stay: retrospective comparative analysis. *BMJ*. 2011;342, d219. doi: 10.1136/bmj.d219

Little RJ, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing data in clinical trials. *New Engl J Med* 2012;367(14):1355-1360.

Lohr KN. Emerging methods in comparative effectiveness and safety: symposium overview and summary. *Med Care*. 2007;45(Suppl 2):S5-S8.
Lord SJ, Irwig L, Bossuyt PMM. *Using the Principles of Randomized Controlled Trial Design to Guide Test Evaluation*. Rockville, MD: Agency for Healthcare Research and Quality; 2009. Medical Tests–White paper series.

Macleod MR, Michie S, Roberts I, et al. Biomedical research: increasing value, reducing waste. *Lancet*. 2014;383(9912):101-104.

Meltzer DO, Hoomans T, Chung JW, Basu A. Minimal modeling approaches to value of information analysis for health research. *Med Decis Making*. 2011;31(6):E1-E22.

Möhler R, Köpke S, Meyer G. Criteria for reporting the development and evaluation of complex interventions in healthcare: revised guideline (CReDECI 2). *Trials*. 2015;16(1):204.

Molnar FJ, Man-Son-Hing M, Hutton B, Fergusson DA. Have last-observation-carried-forward analyses caused us to favour more toxic dementia therapies over less toxic alternatives? A systematic review. *Open Med*. 2009;3(2):e31-e50.

Moore GF, Audrey S, Barker M, et al. Process evaluation of complex interventions: Medical Research Council guidance. *BMJ*. 2015;350:h1258.

Mullins CD, Abdulhalim AM, Lavallee DC. Continuous patient engagement in comparative effectiveness research. *JAMA*. 2012;307(15):1587-1588.

Mullins CD, Barnet B, dosReis S, Kauffman KS, Onukwugha E. Integrating patients' voices in study design elements with a focus on hard-to-reach populations. White paper. Retrieved from http://www.pcori.org/assets/pdfs/ Integrating%20 Patients%20Voices.pdf. Published 2012.

Murray DM. *Design and Analysis of Group-Randomized Trials*. New York, NY: Oxford University Press; 1998.

Muss HB, Berry DA, Cirrincione CT, et al. Adjuvant chemotherapy in older women with early-stage breast cancer. *New Engl J Med*. 2009;360:2055-2065.

National Research Council. *The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials*. Washington, DC: The National Academies Press; 2010.

Organisation for Economic Co-operation and Development. OECD guidelines on the protection of privacy and transborder flows of personal data. Retrieved from http://www.oecd.org/ document/18/0%2C3746%2Cen_2649_34223_1815186_1_1_1_1%2C00.html. Published 2013.

Patient-Centered Outcomes Research Institute (PCORI) Methodology Committee. The PCORI Methodology Report. 2013.

Patient-Centered Outcomes Research Institute (PCORI) Methodology Committee. The PCORI Methodology Report. 2017.

Patient-Centered Outcomes Research Institute (PCORI). Category 5: Standards for heterogeneity of treatment effects. In PCORI Methodology Standards Academic Curriculum. Retrieved from http://www.pcori.org/research-results/ research-methodology/methodology-standards-academic-curriculum. Published 2016.

Patient-Centered Outcomes Research Institute (PCORI). Policy for data management and data sharing. Retrieved from https://www.pcori.org/sites/default/files/PCORI-Policy-for-Data-Management-and-Data-Sharing.pdf. Published 2018.

Petersen ML., Van der Laan MJ. Causal models and learning from data: integrating causal modeling and statistical estimation. *Epidemiology*. 2014;25(3):418-426.

Peterson K, Floyd N, Ferguson L, Christensen V, Helfand M. User survey finds rapid evidence reviews increased uptake of evidence by Veterans Health Administration leadership to inform fast-paced health-system decision-making. *Syst Rev*. 2016;5(1):132.

Pinto DS, Frederick PD, Chakrabarti AK, Kirtane AJ, Ullman E, Dejam A. Benefit of transferring ST-segment-elevation myocardial infarction patients for percutaneous coronary intervention compared with administration of onsite fibrinolytic declines as delays increase. *Circulation*. 2011;124(23):2512-2521.

Platt R, Takvorian SU, Septimus E, et al. Cluster randomized trials in comparative effectiveness research: randomizing hospitals to test methods for prevention of healthcare-associated infections. *Med Care*. 2010;48(6 Suppl):S52-S57.

Pronovost P, Needham D, Berenholtz S, et al. An intervention to decrease catheter-related bloodstream infections in the ICU. *New Engl J Med*. 2006;355:2725-2732.

Raine R, Fitzpatrick R, Barratt H, et al. Challenges, solutions and future directions in the evaluation of service innovations in health care and public health. Southampton, UK: NIHR Journals Library. 2016.

Ray WA. Evaluating medication effects outside of clinical trials: new user designs. *Am J Epidemiol*. 2003;158(9):915-920.

Richardson WS, Wilson MC, Nishikawa J, Hayward RSA. The well-built clinical question: a key to evidence-based decisions. *ACP Journal Club*. 1995;123:A12.

Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. *J Am Stat Assoc*. 1984;79(387):516-524.

Rothwell PM, Fowkes FG, Belch JF, Ogawa H, Warlow CP, Meade TW. Effect of daily aspirin on long-term risk of death due to cancer: analysis of individual patient data from randomised trials. *Lancet*. 2011;377(9759):31-41.

Rothwell PM, Wilson M, Price JF, Belch JF, Meade TW, Mehta Z. Effect of daily aspirin on risk of cancer metastasis: a study of incident cancers during randomised controlled trials. *Lancet*. 2012;379(9826):1591-1601.

Schisterman EF, Cole SR, Platt RW. Overadjustment bias and unnecessary adjustment in epidemiologic studies. *Epidemiology*. 2009;20(4):488-495.

Schneeweiss S, Seeger JD, Smith SR. Methods for developing and analyzing clinically rich data for patient-centered outcomes research: an overview. *Pharmacoepidemiol Drug Saf*. 2012;21(S2):1-5.

Shadish WR, Cook TD, Campbell DT. *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston, MA: Wadsworth Cengage Learning; 2002.

Sheridan S, Schrandt S, Forsythe L, Hilliard TS, Paez KA. The PCORI engagement rubric: promising practices for partnering in research. *Ann Fam Med*. 2017;15(2):165-170.

Smith AJ, Dieppe P, Vernon K, Porter M, Blom AW. Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales. *Lancet*. 2012;379(9822):1199-1204.

Sox HC, Goodman SN. The methods of comparative effectiveness research. *Annu Rev Public Health.* 2012;33:425-445.

Staniszewska S, Brett J, Mockford C, Barber R. The GRIPP checklist: strengthening the quality of patient and public involvement reporting in research. *Int J Technol Assess Health Care*. 2011;27(4):391-399.

Stirman, S.W., Miller, C.J., Toder, K. and Calloway, A., Development of a framework and coding system for modifications and adaptations of evidence-based interventions. *Implementation Science*, 2013;8(1):65.

Thoegersen JL. Examination of federal data management plan guidelines. *J eSci Librarianship*. 2015;4(1):e1072.

US Food and Drug Administration. Guidance for industry: adaptive design clinical trials for drugs and biologics. Retrieved from https://www.fda.gov/downloads/DrugsGuidanceComplianceRegulatoryInformation/Guidances/ UCM201790.pdf. Published 2010a.

US Food and Drug Administration. Guidance for industry and FDA staff: guidance for the use of Bayesian statistics in medical device clinical trials. Retrieved from https://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf. Published 2010b.

US Food and Drug Administration. Clinical outcome assessment (COA): glossary of terms. Retrieved from https://www.fda.gov/drugs/developmentapprovalprocess/drugdevelopmenttoolsqualificationprogram/ucm370262.htm. Published 2015.

Velentgas P, Dreyer NA, Wu AW. Outcome definition and measurement. In: Velentgas P, Dreyer NA, Nourjah P, et al., eds. *Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide*. Rockville, MD: Agency for Healthcare Research and Quality; 2013. Retrieved from https://www.ncbi.nlm.nih.gov/books/NBK126186/.

Verma V, Le T. An analysis of sampling errors for the demographic and health surveys. *Int Stat Review*. 1996;64:265-294

Yordanov Y, Dechartres A, Porcher R, Boutron I, Altman DG, Ravaud P. Avoidable waste of research related to inadequate methods in clinical trials. *BMJ*. 2015;350:h809.

APPENDIX E: CONTRIBUTORS

METHODOLOGY REPORT (2018)

Editors
David Hickam, Emily Evans, Steven Goodman, and Robin Newhouse

Acknowledgments
This revision of the PCORI Methodology Report reflects the insights and efforts of many people. The editors and the Methodology Committee wish to thank these individuals, as well as the PCORI Board of Governors, for their support of the development of the updated methodology standards and this revised report.

PCORI staff served as key members of the topic-focused workgroups that developed the six new standards and accompanying text in the report. These include Rebecca Barasky, Surair Bashir, Laura Esmail, Jason Gerson, and Jamie Trotter.

The updated standards and revised report also benefited from key contributions of external experts, including Peter Craig, Richard Emsley, Robert Kaplan, Laura Leviton, Lucy Savitz, and Meera Viswanathan.

Staff from PCORI's Communications department were instrumental in supporting the public comment process and the design and production of the report, including Marla Bolotsky, Santosh Rao, and Blake Whitney.

METHODOLOGY STORIES AND EXAMPLES (2013)

This part of the Methodology Report is unchanged from the original content published in 2013. The contributors to this content are listed below.

Editorial Team
Writers/Editors: Ayodola Anise, Eric Johnson, Zachary Meisel, Edwin Reid, Lauren Saxton
Designer of Review Materials: Lauren Saxton
Chief Editor for Stories: Mark Helfand

Acknowledgments
We would like to recognize the individuals who supported the development of the stories and examples, including Josh Carlson, Michael Demers, Jacqueline Fridge, Erik 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 contributions to the stories and examples:

- Healthwise (informedhealthdecisions.org) for **Patient Voices: Juli**

- The DIPEx Charity (healthtalk.org) for **Patient Voices: Sarah** (This example is based on research led by the Health Experiences Research Group, Department of Primary Care Health Sciences, University of Oxford.)