Patient-Centered Outcomes Research Institute

Preliminary Draft Methodology Report: "Our Questions, Our Decisions: Standards for Patient-centered Outcomes Research"

Presented by the PCORI Methodology Committee to the PCORI Board of Governors May 10, 2012

Published June 4, 2012

A revised draft will be published for public comment in July 2012.
This draft intended as a resource for PCORI Funding Announcements applicants.

Note: This preliminary draft version of the Methodology Report is provided as a resource for use by applicants for PCORI Funding Announcements issued May 22, 2012, although applications will not be scored on adherence to recommendations in the report as it has not yet had the benefit of public comment. The draft report, in an updated form, will be subject to a formal public comment period, as required by law, starting in July. Comments will be reviewed for use in revising this draft; PCORI's Methodology Committee will deliver a revised version of the report to PCORI's Board of Governors in November, 2012.

PCORI always welcomes public input and individuals can email info@pcori.org at any time with comments. But individuals and organizations might wish to consider withholding formal comment on this draft report pending the start of the official public comment period.

PCORI

The Patient-Centered Outcomes Research Institute (PCORI) was established by Congress through the 2010 Patient Protection and Affordable Care Act, but is by law an independent, non-profit organization. PCORI is governed by a 21-member Board of Governors. It was created to conduct research to provide information about the best available evidence to help patients and their healthcare providers make more informed decisions. PCORI's research is intended to give patients a better understanding of the prevention, treatment and care options available, and the science that supports those options.

(from http://www.pcori.org/about/ & http://www.pcori.org/about/establishment/)

PCORI Methodology Committee

Sherine Gabriel, MD, MSc *(Chair)*, Professor of Medicine and of Epidemiology, William J. and Charles H. Mayo Professor, Mayo Clinic

Sharon-Lise Normand, PhD, *(Vice Chair)*, Professor of Health Care Policy (Biostatistics), Department of Health Care Policy, Harvard Medical School; Professor, Department of Biostatistics, Harvard School of Public Health

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

Ethan Basch, MD, MSc, Associate Attending Physician and Outcomes Scientist, Memorial Sloan-Kettering Cancer Center

Alfred Berg, **MD, MPH**, Professor, Department of Family Medicine, University of Washington School of Medicine

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

Steven Goodman, **MD, MHS, PhD,** Associate Dean for Clinical and Translational Research, Stanford University School of Medicine

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

John Ioannidis, **MD, DSc,** C.F. Rehnborg Chair in Disease Prevention; Director, Stanford Prevention Research Center; Professor of Medicine, Health Research and Policy, Stanford University School of Medicine

Michael Lauer, **MD**, Director, Division of Cardiovascular Sciences at the National Heart, Lung, and Blood Institute

David Meltzer, **MD, PhD,** Chief, Section of Hospital Medicine; Director, Center for Health and the Social Sciences; Chair, Committee on Clinical and Translational Science; Associate

Professor, Department of Medicine, Department of Economics and the Harris School of Public Policy Studies, University of Chicago

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

Robin Newhouse, **PhD, RN**, Chair and Associate Professor, Organizational Systems and Adult Health, University of Maryland School of Nursing

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

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

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

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

The PCORI Methodology Committee wishes to thank the staff, consultants, researchers and others who assisted with the creation of this document.

PCORI Staff

Joe Selby, MD, MPH, *Executive Director*

Anne Beal, MD, MPH, *Chief Operating Officer*

Lori Frank, PhD

Rachael Fleurence, PhD

Gail Shearer, MPP

Bill Silberg

Melissa Stern, MBA

Editing Team

Andrew Holtz, MPH

Heidi D. Nelson, MD, MPH

Ed Reid, MS, MAT

Annette Totten, PhD

Interim Researchers

Tim Carey, MD, MPH

Justine Siedenfeld

Crystal Smith-Spangler, MD

Deloitte Consulting LLC

Constanze Coon, PhD

Anna Cottone

Miliane Ly

Rayneisha Watson, MS, MBA

Responsibility for the final content of this report rests entirely with the Methodology Committee of PCORI.

Suggested citation: Pre-public comment draft report of the Patient-Centered Outcomes Research Institute (PCORI) Methodology Committee. Presented to the PCORI Board of Governors May 10, 2012 and published online June 4, 2012, at www.pcori.org as a resource for applicants for PCORI Funding Announcements issued May 22, 2012. Not intended to be cited for other purposes. To be subject to public comment period starting in July 2012 and revised thereafter.

Forward

Americans today have more healthcare options than ever before: choices between medicine or surgery, between radiation or chemotherapy, between traditional or alternative therapies, and even between treatment or no treatment at all. But sorting out the wheat from the chaff, ensuring that the information used to make those difficult decisions is truly trustworthy can be extremely challenging. The mandate for the Methodology Committee of PCORI is to respond to that challenge by defining methodological standards and a translation table to guide healthcare stakeholders towards the best methods for patient-centered outcomes research (PCOR). Better methods will produce trusted information and lead to better healthcare decisions, and ultimately to better health. With vision, integrity, intellect, and wisdom, and with the support of dozens of scientists from around the U.S., the members of the Methodology Committee have produced the first PCORI Methodology Report. This landmark document (which will continue to be revised and improved over time) is the necessary catalyst for scientifically rigorous, patient-centered outcomes research that can inform decision-making.

Sherine Gabriel, *Chair*
Sharon-Lise Normand, *Vice-Chair*
Methodology Committee of the
Patient-Centered Outcomes Research Institute

Contents

Tables

Figures

Boxes

Appendices

Executive Summary

In this first report, the Methodology Committee puts forward 60 standards to guide patient-centered outcomes research (Appendix A). The initial range of topics was chosen to reflect areas in which the Committee believed that there were either substantial deficiencies or inconsistencies in how the methods were applied in practice, or for which there was specialized knowledge in how best to conduct research that had not been effectively disseminated.(1-3)

This list should be seen in terms of both its overall scope as well as the particular standards that may be most applicable to a particular research project or stage of research. Background, rationale, current practice, recommendations, and other supporting material that help to explain these standards and place them in the proper contexts are included in subsequent chapters. The number of standards within a group should not be taken as any indication of the importance of a topic area.

The report satisfies a requirement of the Patient Protection and Affordable Care Act (PPACA) of 2010, which also established the Patient-Centered Outcomes Research Institute (PCORI) and its Methodology Committee. The legislation states in part that the "purpose of the Institute is to assist patients, clinicians, purchasers, and policy-makers in making informed health decisions by advancing the quality and relevance of evidence concerning the manner in which diseases, disorders, and other health conditions can effectively and appropriately be prevented, diagnosed, treated, monitored, and managed…"(Appendix E-2).

We have interpreted this legislation to mean that the PCORI Methodology Committee's role is to advise PCORI on the best methods for prioritizing, establishing, and carrying out its research agenda by improving the conduct of PCOR, supporting infrastructure for patient-centered outcomes research (PCOR), and furthering the implementation of PCOR to help patients receive optimal outcomes. This first Methodology Committee report focuses on the patient's perspective. Future reports will address clinicians', healthcare purchasers' and policy-makers' perspectives in more detail.

The list of methodological standards for patient-centered outcomes research at the heart of this report is a milestone, but not a destination. Indeed, the legislation establishing PCORI and its Methodology Committee directs that these standards shall be periodically updated (Appendix E-5). Over time, these reports, standards, translation tables, and public engagement forums are expected

to produce better research methods, which in turn will provide information of benefit to all stakeholders—researchers planning an investigation, policy-makers weighing the value of healthcare interventions, and patients and their caregivers facing health decisions. We encourage public comments about the report, the standards, and the translation table and hope to refine the work based on this feedback.

Chapter 1. Introduction

Patient-Centered Outcomes Research

As each of us travels the path of healthy to sick, person to patient, we are faced with a series of choices. Should we have an operation or stick with the pills for our heartburn? Should we take that screening test or skip it and hope this is not the year that we develop cancer? Is it time to bring my feverish child to the doctor or can I wait another day? Each of these decisions, and all the other decisions we make over the course of a lifetime, accounts for much of our personal experience with the healthcare system. The remainder of that experience relates to living with the consequences of the decisions and choices we make. Whether we want to make these decisions, whether we choose nothing instead of something or ask others to make the choices for us, we are still making decisions. Some of us make these decisions after doing our own research at the library or on the internet, while others count on the advice of someone they trust. No matter how we make these decisions, we can all agree on one thing: we all deserve the best possible information to make these decisions. Unfortunately, when it comes to health and healthcare choices, despite the best intentions of the research community far too often the information available isn't good enough.

The information we need to make decisions most often comes from clinical research. Clinical research includes investigations undertaken by scientists, who decide which questions to ask, what approaches to take, how to perform the work, how to interpret the results, and then ultimately how to disseminate the findings through scientific journals or other means. The last 75 years of clinical research has been marked by phenomenal advances in knowledge about the causes of disease and their treatments. Our nation's public and private research funding organizations have helped transform modern medicine, influence the daily healthcare of all of us, and have contributed to unprecedented health and well being of our country. While these successes are all around us, from the perspective of many patients facing health decisions, this research process often misses the mark. Sometimes the research is performed on people who are so different from us that we can't interpret the results. It includes subjects of different ages, sex, race, and without the complexity of conditions that we have. It sometimes involves treatment in care settings not enough like ours—-sophisticated research centers rather than places more like the communities in which we live. It sometimes focuses on choices that don't apply enough to us—expensive treatments that we have to drive hundreds of miles to receive or that we might need to pay for out of pocket … if we have the

money or time. It sometimes deals in outcomes we don't always think are that important—whether or not our blood tests are getting better instead of whether we feel better. For a lot of us, this gap between the information we need and the information we get from research leaves us without the kind of useful information we need to make healthcare decisions. We are often left frustrated by the information we have.

The Patient Protection and Affordable Care Act (PPACA) of 2010 created the Patient-Centered Outcomes Research Institute (PCORI) to support research that can produce the type of information people and their caregivers need when they face a healthcare decision (Appendix E). The purpose of PCORI is to provide the most reliable, relevant, and useful health-related evidence for decision-makers, especially for patients and caregivers. In 2012, the Methodology Committee and the PCORI Board approved a working definition that reflects this perspective:

Patient-centered outcomes research (PCOR) helps people and their caregivers communicate and make informed healthcare decisions, allowing their voices to be heard in assessing the value of healthcare options. This research answers patient-centered questions such as:

1. "Given my personal characteristics, conditions, and preferences, what should I expect will happen to me?"

2. "What are my options and what are the potential benefits and harms of those options?"

3. "What can I do to improve the outcomes that are most important to me?"

What are healthcare interventions?
Treatments, tests, and any other strategies used in the prevention, diagnosis, treatment, and management of illness or injury. The legislation lists them (Appendix E-1):

- protocols for treatment
- care management and delivery procedures
- medical devices
- diagnostic tools
- pharmaceuticals, including drugs and biologicals
- integrative health practices
- any other strategies or items being used in the treatment, management, and diagnosis of, or prevention of illness or injury in, individuals.

4. "How can clinicians and the care delivery systems they work in help me make the best decisions about my health and healthcare?"

To answer these questions, patient-centered outcomes research:

- Assesses the benefits and harms of preventive, diagnostic, therapeutic, palliative, or health delivery system interventions (see sidebar) to inform decision-making, highlighting comparisons and outcomes that matter to people;
- Is inclusive of an individual's preferences, autonomy, and needs, focusing on outcomes that people notice and care about such as survival, function, symptoms, and health-related quality of life;
- Incorporates a wide variety of settings and diversity of participants to address individual differences and barriers to implementation and dissemination; and
- Investigates (or may investigate) optimizing outcomes while addressing burden to individuals, availability of services, technology, and personnel, and other stakeholder perspectives.

To better connect the results of research to the needs of people and their caregivers, PCORI decided they should be involved in defining the basic questions to be asked, how studies are designed and conducted, and ultimately how research results are interpreted and communicated to the people who can use them.

By establishing the definition of and standards for PCOR, funding PCOR investigations, building infrastructure to support this approach to studying health and healthcare, and through other means, PCORI aims to promote and catalyze the development of evidence that is relevant at the time and place that people and their caregivers make health decisions.

The legislation for PCORI established a 17-member Methodology Committee selected by the General Accounting Office "to develop and improve the science and methods of comparative clinical effectiveness research," and to report on methodological standards for research (Appendix E-5). In order for PCORI to fulfill its mandate "to assist patients, clinicians, purchasers, and policy-makers in making informed health decisions," critical problems related to the evidence used to support health decision-making need to be addressed (Appendix E-2). We have interpreted this legislation to mean that the PCORI Methodology Committee's role is to advise PCORI on the best methods for promoting its agenda—in prioritizing research, establishing a specific research project agenda, reviewing research proposals, supporting infrastructure for PCOR, improving the conduct

of PCOR, and furthering the implementation of PCOR to help patients receive optimal outcomes. The legislation requires the Committee to create standards and recommended actions and issue reports and a translation table to help PCORI fund high impact patient-centered research and accomplish the organization's mission.

Problems that PCORI Hopes to Address

The founding legislation for PCORI is based on the premise that there is an opportunity for research and research funding agencies to better support decision-making by patients, caregivers, clinicians and policy-makers. To address this opportunity we thought it best to begin by identifying the key problems in existing research from the perspectives of the patient and of these other stakeholders. This itemization highlights the effect these problems have on healthcare decision-makers and calls out the ways PCORI can address them.

- *A relevance problem related to questions and endpoints*—Researchers often choose questions and endpoints that they consider to be of interest and importance. Sometimes these are not the questions and endpoints most relevant to the people affected. Researchers often pursue these types of endpoints because assessing them is more feasible, less expensive, and because the time required to look at the ultimate outcome may take too long to inform decision making in the short-term. For example, a researcher might study whether or not a stent placed in an artery to improve blood flow is open 30 days after placement. While that question may be reasonable to some, the question that patients want answered is whether symptoms and function, like pain and the ability to walk, are better after the stent was placed. They want to know whether a stent is the best way to achieve the benefits they are hoping for. Measuring whether the stent is open at 30 days may be easier than measuring walking and quality of life at a year, but just because it's easier to assess doesn't mean studying that endpoint will help patients decide if a stent is right for them.

  Research can also lack relevance when the big questions patients ask are divided up into pieces that researchers are more able to tackle. While researchers hope that eventually all the pieces get addressed, failing to connect research programs to clinical decisions means that too often not all the pieces are addressed, leaving decision-makers with an incomplete, frustrating patchwork of information.

A "patients like me" problem—Research sometimes focuses on patients with a narrow set of characteristics and conditions. Often there are practical purposes for this—it takes a much larger study to account for differences between patients, and the bigger the study, the greater the cost. Sometimes there are scientific purposes—narrowing the number of variables in a trial of a new drug makes it more likely that any effects are due to the drug and not something else. Sometimes researchers want to include patients with broader characteristics but struggle to accomplish that goal. They often have trouble recruiting study participants who represent the full spectrum of patients. Many people are reluctant to participate in research because they lack trust in researchers, don't have the time, or are even scared by the concept of being involved in research. For whatever reason, if research doesn't account for "our" characteristics, then the results aren't as likely to be relevant to us. What people want is information that takes into account all their unique characteristics and conditions. Personalizing research results is becoming ever more challenging as we learn more about the genetic variation that makes each of us different from the "average" patient.

It was always hard to find time to catch up with the medical literature. The doctor had been looking forward to reading a recently released report on a randomized trial for a new rheumatoid arthritis drug. He was curious if it might be a good fit for some of his patients.

When he got to the methods section of the article, however, he was disappointed. For one, the inclusion criteria were narrow. It seemed that this study excluded participants with additional issues that could complicate their treatment, just the sort of issues faced by most of his patients, such as high blood pressure, heart problems, or liver disorders.

The outcomes reported in the trial—the number of joints affected, and how well they function—were improved for 75% of the participants. However, he knew from past trials that drugs are often more successful in a trial setting than in clinical practice. The narrow criteria for patient selection could make it more likely that each participant in the study would respond to the drug in a similar way. When used in a more diverse group of patients, the results might not be as good.

The doctor pondered the results of the study. It was true that many of his patients did not meet the study criteria, but this did not necessarily mean this drug would be useless for them; many treatments tested in randomized trials with designs like this have proven to be effective in a wide variety of people. He decided that the jury was still out on this treatment, and while he would mention it to patients who were not doing well on other treatments, he would wait for future studies to test this treatment on a wider segment of the population with outcomes that could be applied directly in clinical practice and more information on side effects before he would tell all his patients about it.

Sources: (4, 5)

- *A trust problem*—There are serious concerns that research being proposed, reviewed, funded, and performed is not independent enough of those with a financial or professional interest in the results. Distrust is also the result when the aims of researchers and the people they want to study are not aligned. Typically, decisions about whom to include in a study, how to deliver treatments, how long the study should continue, and what outcomes to measure are most often made by funders and researchers, whose intellectual or financial bias can influences these and other aspects of a study's design. For example, a specialist who has developed a new procedure and is the main researcher of that procedure may be more interested in the positive outcomes of treatment but less interested in the risks. A pharmaceutical company that conducts studies of its own products will normally choose whom to study, what to compare, and what outcomes to measure so as to support an application for regulatory approval and ultimately success in the marketplace. As a result, the apparent efficacy or safety of a medical treatment may depend on who is studying it and what they have to gain or lose by the results. Sometimes, rather than performing a comparison to the best alternative treatment to

see which is better, investigators may choose, and sometimes regulatory agencies encourage, a comparison to a placebo, or an inferior alternative. Sometimes investigators are less inclined to publish results when the study shows no difference. Failure to fully publish the results of research also undermines trust and can create a false impression of the effectiveness and safety of treatments.(6)

Overall, randomized trials of treatments for hypertension (high blood pressure) demonstrate that lowering blood pressure reduces the risk of stroke, heart attacks, heart failure and other important health outcomes, and that for most people results are as good with older, generic drugs as they are with newer ones. In the 2000's,in some highly publicized observational studies, patients taking one type of drug—the calcium channel blockers— had higher rates of heart attacks than those taking other drugs. The use of this type of drug fell rapidly. But did these observational studies get the right answer for everyone?

In 2008, a large controlled trial studied patients who still had poor control of blood pressure, despite being on two medications. More than 11,000 patients were randomized to combination treatment with an ACE inhibitor (benazepril) and either a diuretic (hydrochlorothiazide) or calcium channel blocker (amlodipine). After an average follow-up of three years, the patients who received the calcium channel blocker had a 2 percent lower absolute risk (about 20 percent lower relative risk) of cardiovascular illness or death.

While there was debate about the meaning of the results (with some noting that diuretics were under-dosed, that a different diuretic might have worked better, and that the study was funded and the data analyzed by a company that sells the calcium channel blocker), there was broad acceptance that this trial provided better evidence than the observational studies for some high-risk patients and provided useful guidance to physicians trying to choose from the multitude of anti-hypertensive drugs available.

Sources: (8-12)

- *A quality problem*—The phenomenal volume of research obscures high-quality studies (that employ the right methods in the right way) in a clutter of reports that fail to provide results useful to those making health decisions. Not infrequently, comprehensive reviews of research about a clinical problem find that many studies re-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.(7)
- *A user-friendliness problem*—The public wants and needs research results clearly connected to their health decisions. Many patients and most of their caregivers also want information presented so that they can understand the strengths and limitations of a given research design. We are barraged with healthcare information: conversations about a friend or family member's healthcare experience; claims made by promoters of a certain procedure, medicine, or approach; and media accounts of purported medical breakthroughs. Certain segments of the public want to engage in PCOR, but believe the methods being used are not responsive to their needs. Some want to be involved in debates about methodological alternatives and tradeoffs in traditional study design so that they can shape the way healthcare questions are addressed. The public and those interested in being part of the research process need a way to understand how different methods for conducting, evaluating, and implementing healthcare research are likely to produce useful information.

- *A priority problem*—When it comes to health and healthcare interventions, there are so many important questions, and with limited research dollars we need to prioritize. We need a set of methods to guide which questions the system should tackle first and a way to keep patients central in that prioritization process. Too often this ranking has been done out of

public view through a process subject to political and economic forces that lacks a coherent strategy. Applying methodological standards to set priorities will produce "winners and losers"—some topics will receive more funding support sooner than others. Like any funding agency, PCORI cannot be all things to all people. If it places the highest priority on research questions that will offer the greatest good for the greatest number of people, then it might be thought to signal a focus on common conditions or majority populations to the dismay of advocates for those with rare but serious diseases or minority populations. Conversely, focusing on less common conditions or smaller populations with the loudest advocates could drain research funding without producing adequate progress against widespread health threats. PCORI's Methodology Committee promotes a thoughtful and reasoned approach to prioritization using a suite of techniques that includes value of information analysis and incorporates all the domains of relevance when deciding what to fund.

- *An implementation problem*—It takes too long to disseminate research results to providers, patients, and other stakeholders so that discoveries can be applied to health decisions and support the most effective healthcare practice. PCORI hopes to develop and apply optimal methods for encouraging the adoption of what we know into what we do.

- *A record-keeping problem*—While millions of American doctor and hospital and other health visits are now recorded in electronic medical records (EMR), the vast majority of those electronic data cannot be used for research. This enormous potential of EMR to answer significant PCOR questions remains largely untapped. PCORI is just beginning to explore how to leverage its investment in this arena to achieve substantive improvements in these systems to permit their use as a research tool, answering questions of meaning to patients.

Addressing these problems is the purpose of these and future Methodology Committee standards, recommendations to the PCORI Board of Governors, reports, and translation tables.

An Evolving Document

PCORI aims to incorporate the standards and recommendations discussed in this report into the funding process and will encourage their adoption by the broader scientific community. The next three years of Methodology Committee work will be a continual process of reconsidering, refining, and widening the scope of the standards to include the full spectrum of PCOR questions and approaches. Similarly, the translation table within this report will be expanded over time to include more examples, methodological issues, and approaches.

The list of methodological standards for patient-centered outcomes research at the heart of this report is a milestone, but not a destination. Indeed, the legislation establishing PCORI and its Methodology Committee directs that these standards shall be periodically updated (Appendix E-5). Over time, these reports, standards, translation tables, and public engagement forums are expected to produce better research methods, which in turn will provide information of benefit to all stakeholders—researchers planning an investigation, policy-makers weighing the value of healthcare interventions, and patients and their caregivers facing health decisions. We encourage public comments about the report, the standards, and the translation table and hope to refine the work based on this feedback.

The legislation establishing PCORI conveys a sense of urgency related to setting methodological standards and establishing national priorities. Federal funding also helped support related activities related to setting priorities including ; 1) an Institute of Medicine study to develop standards for conducting systematic reviews of comparative clinical effectiveness,(13) a crucial tool for setting priorities; 2) a report on Initial National Priorities for Comparative Effectiveness, conducted by an Institute of Medicine panel(14) that included patient and consumer representatives and that applied criteria for prioritization similar to those specified in the Affordable Care Act; and 3) funding for an initial program of comparative clinical effectiveness research through the Agency for Healthcare Research and Quality (AHRQ) and the National Institutes of Health (NIH). The standards proposed in this report can help PCORI use this previous work effectively and efficiently to prioritize and establish a specific research project agenda.

Contents of this Report

Chapter 2 of this report focuses on how the Committee approached this core task of developing methodological standards, and Chapter 3 provides an overview of the standards and actions we recommend to make them effective. We then describe the rationale for standards for patient-centeredness (Chapter 4); prioritizing topics for research (Chapter 5); choosing a study design (Chapter 6); and for designing, conducting, and reporting research (Chapter 7-8). These chapters also highlight gaps in the evidence that should be addressed by PCORI's program of methodological research. Finally, we describe priority areas for the Methodology Committee over the next 2-3 years (Chapter 9).

Chapter 2. How the Methodology Committee Developed the Recommended Standards

The Methodology Committee's Approach to the Authorizing Legislation

The Methodology Committee began its work by defining terms and considering the various approaches that might be taken in responding to requirements in the authorizing legislation. Three specific activities were named: the creation of a "translation table," the proposal of methodological standards for research, and the proposal of activities for enforcing the methodological standards.

Translation table. The legislation specifies that the Methodology Committee must develop or devise "a translation table that is designed to provide guidance and act as a reference for the Board to determine research methods that are most likely to address each specific research question" (Appendix E-5). In Chapter 6, the Committee proposes a general framework for the translation table, with several of the research domains presented as examples of what a final product might look like. It is expected that the translation table will eventually include versions appropriate for use by different stakeholder groups, and that the Committee will also define ways to further develop the concept and evaluate its usefulness.

Methodological standards for research. The language of the legislation around the establishment of methodological standards has several salient features that outline the Congressional expectations for the committee and its work. First, the legislation defines expectations with regard to the *process* by which the committee should proceed, stipulating that the establishment of methodological standards should build on existing work, and that the process should be ongoing, scientifically based, and inclusive. Second, the legislation specifies both specific and general *content* areas that the proposed standards should address, including internal validity, generalizability, feasibility, and timeliness of research. Third, the committee is to provide specific criteria for health outcome measures, risk adjustment, and other relevant aspects of research and assessment with respect to design of research. Fourth, standards must be recommended by which patient subpopulations can be accounted for and evaluated. Finally the *scope* of the committee's work is to include "each of the major categories of comparative clinical effectiveness research methods" (Appendix E-5), which are listed as:

- "Systematic reviews and assessments of existing and future research and evidence including original research conducted subsequent to the date of the enactment of this section.
- Primary research, such as randomized clinical trials, molecularly informed trials, and observational studies.
- Any other methodologies recommended by the methodology committee established under paragraph (6) that are adopted by the Board under paragraph (9)" (Appendix E-3).

Recommended actions to comply with methodological standards. The enabling legislation instructs the Committee not only to propose methodological standards but also to recommend actions necessary to comply with them:

"The methodology committee shall submit reports to the Board….Reports shall contain recommendations for the Institute to adopt methodological standards…as well as other actions deemed necessary to comply with such methodological standards" (Appendix E-6).

The Methodology Committee and PCORI Board and staff are developing a coordinated approach to promote the uptake of PCORI methods standards. This includes engaging all stakeholders who might use the standards, creating reporting and surveillance opportunities, and developing enforcement functions over time.

Define Methodological Standards

A comprehensive resource specifying standards for research in all relevant aspects of comparative effectiveness research (CER) with a patient-centered focus would be a vast undertaking, and the Committee's initial work in this space was defined more narrowly. The Committee presents here a first installment of what will be an ongoing task, both in completing the inventory of recommended standards and in periodically reviewing and updating them. The Committee identified general areas for its initial focus that would allow PCORI to address its first round of funding announcements and priority development. Small working groups were formed which explored patient-centeredness, research prioritization, and research methods. A fourth group coordinated communication and prepared draft report components. The goal was to propose methodological standards in important research domains that are representative of research issues in CER and that will eventually be covered more comprehensively.

Building on the work of the Institute of Medicine,(13) the Committee defined a standard as follows:

- A process, action, or procedure for performing PCOR that is deemed essential to producing scientifically valid, transparent, and reproducible results. A standard may be supported by scientific evidence, reasonable expectation that the standard helps achieve the anticipated level of quality in PCOR, or by broad acceptance of the practice in PCOR.
- The recommendation is actionable, feasible, and implementable.
- Proposed standards are intended for use by the PCORI Board, in PCORI policies and procedures, and by PCORI researchers.

Select and Assess Proposed Standards

Committee working groups developed provisional lists of major research methods questions and chose 129 for focused review for this first methods report, based on considerations presented in Box 2.1, below. Contractors were secured to assist the Committee in developing materials for each topic (see Appendix G). (Full reports are available at http://www.pcori.org/what-we-do/methodology/. Citations were not carried forward into this report to increase readability, but are fully documented in the contractor reports posted on the PCORI Web site.) In addition to full reports, contractors summarized key information regarding each proposed standard in a template format to assist Committee members in making direct comparisons between proposed standards (see Table 2.1) based broadly on criteria derived from AGREE, an international project developing guidelines for the appraisal of research and evaluation (Box 2.1). Four criteria were deemed especially important: contribution to patient-centeredness, contribution to transparency, contribution to scientific rigor, and empirical evidence/theoretical basis. The workgroups of the Committee held workshops at which contractors presented their findings and recommendations for discussion with Committee members and invited experts, including patient representatives. In addition to reports from contractors, the Committee relied on its members to bring their experience, knowledge, and additional publications and documents to the table. Committee workgroups reduced the original list of 129 methods questions to 88 potential standards by eliminating those clearly not yet ready to be considered standards and by removing those that were out of scope or redundant.

Materials were then developed around the 88 proposed standards that Committee members reviewed in depth. A preliminary poll was taken that informed a meeting attended in person or by teleconference by all Committee members at which standards were discussed. Formal votes were taken, requiring approval by a minimum of two-thirds of members (12 of 17) for a standard to be

recommended. In the final voting, Committee members were asked to consider each standard as a minimum requirement for PCORI. Thus some proposed standards, while perhaps finding agreement that they may be good practice, were not recommended as standards. The Committee agreed that there would be no formal "minority" statements for those standards not approved or for approved standards where approval was not unanimous, although some of the complexities in the decisions are presented in the detailed discussions that follow presentation of the standards in the chapters that follow.

Box 2.1. Modified AGREE Criteria

- The purpose of the work is to define methodological standards for CER.
- The people to whom the standards apply are described.
- The application of the standards to CER is clear.
- The standards were developed by a relevant professional group.
- Patient's views and preferences were sought.
- Stakeholders were involved in the development of the standards.
- A systematic process was used to generate recommendations.
- Details of the systematic process used to generate recommendations are provided.
- There is an explicit link between the rational for the standard and the recommended standard itself.
- The standards were externally reviewed before publication.
- The recommendations are specific and unambiguous.
- Key recommendations are clear.
- The standards are editorially independent from the funding body.
- Conflicts of interest have been recorded.

Modified from AGREE Instrument Available at: http://www.agreetrust.org/

Table 2.1. Committee Criteria for Adopting Proposed Standards

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

Finally, in addition to the recommended standards, the Committee reached a number of consensus conclusions regarding recommended actions, directed to the PCORI Board of Governors. These statements appear throughout the report and are collected in Appendix B.

As the basis for developing a translation table and framework, Committee staff searched MEDLINE, Scopus, and the AHRQ Scientific Resource Center Methods Database (http://www.citeulike.org/user/SRCMethodsLibrary) for articles that use the term "translation table." The results were reviewed by Committee members. A Request for Information was also used to gather stakeholder input regarding the translation table (see Appendix D-1, D-2).

Chapter 3. Overview of the Standards

Background

There are few more important aims of medical research than to conduct studies that provide accurate estimates of benefit or harm. But no estimate can be perfectly accurate; there is always some uncertainty around it. Getting that uncertainty correct is as important as the estimate itself. This is particularly important for PCOR, which is specifically designed to inform decision-making. For example, if a study estimates a two-fold increase in a serious side effect due to a drug interaction, decisions based on that information could be very different if the confidence interval around that 2-fold estimate were 0.5 to 4, indicating considerable doubt about the existence of the effect, as opposed to 1.9 to 2.1, indicating that the 2-fold estimate was right on the mark.

However, understanding all the contributors to uncertainty often takes years of development and application of a method, just as it does for any applied technology. Such experience leads to an appreciation of the potential pitfalls in practice and in theory, and the ways to avoid them. In this way, methods evolve and improve over time. But there are no formal mechanisms to guarantee that best practices are adopted. One of the more effective ways is for the sponsors of research to articulate standards for the conduct of research they fund. This was recognized in the PCORI authorizing legislation, which tasked the Methodology Committee with developing and promulgating such standards (Appendix E-5).

The pursuit of medical knowledge has evolved from predominately case studies to clinical trials and observational studies that use increasingly complex designs and analytic methods. Over the past four decades, explicit, formal standards for planning, conducting, and reporting clinical trials were developed for the subset of research studies that were conducted to obtain regulatory approval from the U.S. Food and Drug Administration. These standards, articulated in formal "guidance documents," helped to create a level playing field for regulatory decision-makers and for the companies designing such studies (http://www.fda.gov/RegulatoryInformation/Guidances/ucm122049.htm). The wider range of research studies used to inform clinical decisions has been governed by less formal approaches, such as peer review of research proposals and of reports submitted to scientific journals. Many specialty societies or ad hoc groups of experts have taken it upon themselves to issue methodological

A Thought Experiment

Consider two treatments for rheumatoid arthritis. We'll call them Alpha and Beta. Alpha requires infusions given in a clinic. Beta is a pill taken each morning and evening. Previous studies have shown that each is effective. What researchers want to do now is compare them in a setting as close as possible to typical clinical use. Special outreach to patients in low-income neighborhoods and rural areas, as well as the use of materials and practices suited to people with limited education and those who don't speak or read English well, produce a study population that resembles the broad spectrum of people being treated from rheumatoid arthritis in routine clinical practice.

After randomizing study participants to start on Alpha or Beta, the researchers step back, allowing the patients, their doctors and nurses to make their own healthcare choices, including switching treatments, just as in routine practice. And just as in actual practice, the patients' care is covered by their existing insurance, rather than a special grant.

When the data is analyzed, it appears that Beta, the twice-a-day pill, appears to be more effective than Alpha, the clinic-administered infusion, at relieving pain, improving daily function and other outcomes important to patients.

But before writing up their results, the researchers look for explanations. It wasn't side effects or other biomedical parameters that made the difference. Two themes emerged from patient reports. Some stopped using Alpha because it cost them more out of pocket. Others said that even if they could afford the copays, they couldn't take time off work to get infusions or they had trouble finding transportation to the clinic.

In this example, then, the results do not tell us much about what can work better, because out-of-pocket costs, access to care, insurance, and perhaps even more remote factors (such as the frequency of bus service to the medical center or from there to work) influenced the result.

guidelines in the areas of their expertise. A related comprehensive effort has been ongoing for several decades to develop reporting standards for studies that employ specific designs. These guidelines are currently housed at the Equator network Web site (www.equator-network.org), including widely utilized tools such as CONSORT (for randomized clinical trials), STROBE (for observational studies), and STARD (for diagnostic accuracy studies). However, these guidelines are specifically for reporting, not study conduct, although they can be interpreted as indirect guides to how studies should be executed. In 2008, the Institute of Medicine stated that methodological standards for the conduct of one type of research—systematic reviews—would help decision-makers involved in PCOR "with respect to transparency, minimizing bias and conflict of interest, and clarity of reporting."(15) In 2011, the Institute of Medicine published such standards.(16)

Balancing the Aims of PCOR—A Critical Methodological Challenge

The standards presented in this report attempt to balance several aims of PCORI. One aim is to counter overgeneralization from studies in narrow populations or in highly controlled research settings to the broader range of people and settings in "real life." (17, 18) But another aim, undoubtedly, is to learn more about how well the different treatment choices *can* work, and for whom they work best and are safest. Designing studies so we can be confident that the treatment is producing the results is particularly difficult in clinical effectiveness research. (19) (See the box entitled "A Thought Experiment" for an illustration of this point.) The Committee will return to the issue of balancing these aims in the near future.

Recommended Standards

In this first report, the Methodology Committee presents 60 standards to guide patient-centered outcomes research (Appendix A). The initial range of topics was chosen to reflect areas in which the Committee believed that there were either substantial deficiencies or inconsistencies in how available methods were applied in practice, or for which there was specialized knowledge in how best to conduct research that had not been effectively disseminated.(1-3)

This list should be seen in terms of both its overall scope as well as the particular standards that may be most applicable to a particular research project or stage of research. Background, rationale, current practice, recommendations, and other supporting material that help to explain these standards and place them in the proper contexts are included in subsequent chapters. The number of standards within a group should not be taken as any indication of the importance of a topic area.

The standards offer an approach to align the research agenda with questions that underlie patients' and clinicians' uncertainty about what works best, for whom, under what circumstances. Methodological standards aim to do this by improving the way research questions are selected, formulated and addressed, and findings reported. They can also help prevent the research agenda from employing flawed, out-of-date, or inappropriate methods to answer research questions, and may raise the bar for researchers, publishers, and industry as they try to inform decision-making. Just as standards for premarketing studies helped level the playing field by defining the requirements for decisions about regulatory approval of a new drug or device, standards for PCOR can also benefit medical innovators by providing a common set of expectations about the characteristics of high-quality research.

The potential range of methods for which standards could be developed is substantially larger than what is presented herein and will be the subject of further standards development in future PCORI Methodology Committee work. We do not present standards for every aspect of conducting research. Rather, the standards focus on a few methodologies and issues that the Methodology Committee identified as likely to contribute to rapid improvement in the quality and value of PCOR.

We do not intend our articulation of standards to inhibit methodological innovation, and, in fact, hope to promote such innovation. For this reason, in the areas of research methodology addressed in Chapters 7 and 8, the Committee chose to present "minimal" standards, those practices that few

would dispute are necessary for sound science and that should not inhibit further evolution of the methods. All of the standards presented have scientific justification, either from empirical studies or from theoretical considerations, which typically apply to mathematical methods. Quite a few promote transparency: how to properly communicate, both in study protocols and in published reports, exactly what was planned and what was done.

In other areas, particularly in methods to engage patients in prioritizing and refining research topics, it is not possible to identify evidence-based standards. We believe that standards for engaging patients in each phase of the research process are essential, but lack the evidence to specify which methods for doing so are best. Several of these patient engagement standards call for researchers to describe the methods they propose or employ. Such information will allow for evaluation of the performance of different research methods. As more PCOR is conducted and as more evaluation of patient engagement occurs, more specific standards, including those addressing engagement of stakeholders beyond patients, will be developed.

Scope of the Standards: The Phases of PCOR

The standards address selected topics in four broad phases or categories of activities (Table 3.1):

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

The table lists key activities within each phase and provides selected details for each activity. The list of activities is not intended to be comprehensive, and only selected details are provided. For example, while Standard 4.1.1 (Chapter 4) calls for incorporation of patient representatives in the full range of research phases and activities, this is not explicitly noted in Table 3.1 other than for Phase 3, "How do we carry out and govern the study." Furthermore, the specific phases and activities involved in any given study may vary, e.g., differing for patient-level studies (e.g., of treatments) vs. system-level studies comparing delivery system interventions.

Table 3.1. Phases of PCOR

Phase of PCOR	Details of phase
1. What should we study?	
Identify and define important research questions	• Identify topics, decisions, and questions that are important to patients, caregivers, and other stakeholders • Specify the research questions in a manner highlighting patient-centered outcomes and information needs
Prioritize research questions	• Decide on the importance and priority of topics and questions, taking into account evidence gaps and the value of information
Refine and specify details of research questions	• Specify the population, interventions, comparators, outcomes, timing, and setting (PICOTS) to accurately capture each research question
Develop funding announcements	• Develop and release funding announcements for high-priority questions • Incorporate guidance and standards to ensure alignment with the resulting projects
Conduct peer review and funding decisions	• Incorporate guidance to reviewers to facilitate assessment of investigator responsiveness to patient-centeredness aspects of studies
2. What study designs should we use?	
Create conceptual framework	• Specify hypothesized mechanisms of effect, influences on outcomes, and clinical context of affected care processes to guide data collection and analysis plans
Select research approach and study design	• Select study designs and research approaches (quantitative and/or qualitative) that suit the question, taking into account tradeoffs in the internal validity, generalizability, feasibility, and timeliness of each approach (see Translation Table)
Design data collection and analysis plans	• Select data sources and specify data collection methods most likely to produce valid, useful data • Design analysis plan to optimally answer the research questions
Develop dissemination and related follow-up plans	• Assess the likely value and relevance of projected study findings for each potential stakeholder group (including policy, practice, patient, research) and identify appropriate use of findings • Develop broad outline of dissemination plans and plans for follow-up for each affected stakeholder group and projected use of findings
3. How do we carry out and govern the study?	
Employ strategies to reduce dropouts and maximize validity of data	• Work with representatives of study population to achieve ongoing engagement and support for accurate, complete data
Employ strategies to maximize validity of analysis and interpretation of results	• Work with representatives of study population and affected stakeholder groups to review and discuss study findings and consider alternative explanations

Phase of PCOR	Details of phase
4. How do we enable people to apply the study results?	
Report the results	• Convey the findings of research in a comprehensible manner that is useful to patients and providers in making healthcare decisions • Fully convey findings and discuss considerations specific to certain subpopulations, risk factors, and comorbidities, as appropriate • Discuss limitations of the research • Do not include any data which would violate the privacy of research participants or any confidentiality agreements (Appendix E-7)
Assess study implications	• Assess implications and appropriate "next steps" (follow-up) to study results (e.g., incorporation in systematic reviews; progression to larger, more definitive study; dissemination to policy/practice audiences)
Disseminate	• Disseminate study findings to appropriate audiences based on their identified implications for research, policy, and practice • Provide necessary interpretation and explanation to facilitate stakeholder understanding and appropriate interpretation of findings and implications
Implement	• Plan and conduct activities to facilitate implementation of study results if they are relevant to policy or practice

The phases and activities listed in the table correspond with the scope of work of PCORI. The legislation sets instructions for PCORI to prioritize, establish, and carry out a research agenda, making reference to methodological standards (Appendix E-3). The standards concern aspects relevant to these duties, with a particular focus on the design of research studies (Appendix E-5). The standards presented here both draw from and extend this literature to advise researchers funded by PCORI and others conducting PCOR what is expected of them in the design, conduct, analysis, and reporting of their studies. We hope that these standards, like the methods themselves, will draw comment and evolve over time. The ultimate goal is to produce research that patients, caregivers, and other stakeholders in medical decision can rely on.

We focused most of our attention on the earlier phases, those needed to support PCORI's first specific agenda for comparative clinical effectiveness research. Many of these standards specify what to include in research protocols and reports.

The first set of standards does not cover all of the possible combinations of types of studies and interventions set out in the legislation. Most notably, although systematic reviews of effectiveness play a central role in PCOR, we did not develop standards for them because the Institute of Medicine has recently done so. The Committee plans to review and will consider endorsing some or all of them. Some topics we consider to be high-priority, such as cluster randomized trials, additional

aspects of the evaluation of diagnostic tests, systematic reviews of system interventions, and modeling, will be addressed in future reports.

Transparency in Research: The Context for Implementing the Standards

How PCORI and other funding agencies apply the Committee's standards will help determine whether the perceived problems with research are addressed in an effective, equitable, and transparent fashion. Basic good research practices are a required foundation for the standards for PCOR. Departures from basic good research practices are partially responsible for the mismatch between the quality and relevance of the information research provides and the information patients need to make informed clinical decisions.

One of the most important components of this foundation is a commitment to transparency in research. Transparency is needed to enable researchers to verify research findings. Many of the standards promote transparency by requiring detailed protocols for proposed research and compliance with guidelines for study registration and reporting results. Not only can these requirements help PCORI judge the quality and relevance of proposed research plans, but they also may help protect against practices, such as selective reporting, that can distort or misrepresent research results.

The Methodology Committee recommends that, in the near future, PCORI develop and implement additional policies to encourage research that is transparent and reproducible (see Recommended Standards for Transparency, page 23). Specifically, the Methodology Committee recommends that PCORI develop policies to encourage public registration of all PCORI studies and the sharing of study protocols, statistical code, and data. A standing committee within PCORI should be formed to assess appropriate methods for data sharing and to ensure that proper scientific credit is given to those sharing protocols, code, and data (see Recommended Actions for PCORI).

PCORI has a stake in transparency and reproducibility not only in the research it undertakes, but also in research conducted by other entities. For example, systematic reviews, which are a key type of research for PCORI and which underlie judgments about future research needs, are highly dependent on the degree to which evidence is reported fully and in an unbiased manner. Credible standards for conducting systematic reviews specific to clinical effectiveness recognize that "reporting biases, particularly publication bias and selective reporting of trial outcomes and analyses,

present the greatest obstacle to obtaining a complete collection of relevant information on the effectiveness of health care interventions."(16) An important next step for PCORI is to develop policies that can remove or overcome this obstacle, not only in its own research, but throughout the broader clinical research community.

Standards for Formulating Research Questions

3.1.1 Develop a Formal Study Protocol

The study protocol should include all the elements specified in the standards.

3.1.2 Identify Specific Populations and Health Decision(s) Affected by the Research

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

3.1.3 Identify and Assess Participant Subgroups

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

3.1.4 Select Appropriate Interventions and Comparators

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

3.1.5 Measure Outcomes that People in the Population of Interest Notice and Care About

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

Preparing and publishing a detailed research protocol and asking a well-formulated research question are critical for transparency. Many of the 60 standards specify what should be included in a research protocol. The first requirements are components of a well-formulated research question, often abbreviated "PICOTS":

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

These standards are part of a larger group of general standards that apply across all research on patient-centered outcomes. We return to them throughout this report.

In the next chapter, we propose standards for making the patient's voice the central focus of defining research questions.

Recommended Actions for Transparency

- The Methodology Committee recommends that PCORI develop policies to encourage public registration of all PCORI studies and the sharing of study protocols, statistical code, and data.
- Form a standing committee within PCORI to recommend appropriate methods for data sharing and to ensure that proper scientific credit is given to those sharing protocols, code, and data.
- To speed implementation of standards in funding announcements, peer review, and other internal processes, PCORI staff should develop or have developed templates for the preparation and review of proposals that incorporate the key elements of the standards. Because some standards apply only to certain types of studies, a portfolio of templates applicable to various study designs should be developed.

Chapter 4. Methodological Standards for Patient-Centeredness of Research Proposals and Protocols

Introduction

Patient-centered outcomes research helps people to make informed healthcare decisions by directing research toward questions that are important to patients, measuring outcomes that are noticeable and meaningful to them, and producing results that help them weigh the value of healthcare options given their personal circumstances, conditions, and preferences (see Chapter 1). While *patient* is defined here as any individual with or at risk for a specific health condition, this discussion applies also to caregivers and patient surrogates where appropriate. We take special notice of patients who may be hard to reach because of socioeconomic, geographic, racial, or ethnic barriers or due to physical or cognitive impairments, and we are especially aware of the efforts needed to ensure that researchers engage the full spectrum of individuals and communities.

There are many other individuals involved in any health decisions, often referred to as stakeholders, including clinicians, healthcare system administrators, payers, regulators, and policy-makers. In this report we focus on patients rather than on other health decision stakeholders, who will be the focus of future standards.

Patient-Centeredness Standards

PCOR starts from the vantage point of individuals facing health decisions. Every step of the design, conduct, analysis, and dissemination of PCOR should be directed towards informing health decisions that affect outcomes that are meaningful to a specific group of patients. From the earliest phases of defining a research topic and formulating a study question; then identifying a study population and choosing interventions, comparators, and outcomes to measure; through the conduct of a study and analysis of results; and ultimately to the dissemination of research findings into clinical practice, researchers should ensure PCOR results accurately and effectively inform health decisions important to patients.

Standards for Patient-Centeredness and Engagement

3.1.2 Identify Specific Populations and Health Decision(s) Affected by the Research

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

3.1.5 Measure Outcomes that People in the Population of Interest Notice and Care About

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

4.1.1 Engage Patient Informants, Persons Representative of the Population of Interest, in All Phases of Patient-Centered Outcomes Research (PCOR)

Research proposals should 1) describe how patient informants will be: identified, recruited, and retained; involved in determining the study design and monitoring of its conduct; and involved in dissemination of research results, and 2) state how the research process will follow PCOR principles of trust, transparency, co-learning, respect, and partnership. Patient informants include individuals who have the condition or who are at risk of the condition, and, as relevant, their surrogates or caregivers. At a minimum, patient informants should be engaged in formulating research questions; defining essential characteristics of study participants, comparators, and outcomes; monitoring study conduct and progress; and disseminating results.

Standards for Patient-Centeredness and Engagement (continued)

4.1.2 Identify, Select, Recruit, and Retain Study Participants Representative of the Spectrum of the Population of Interest Facing the Health Decision of Interest and Ensure that Data Are Collected Thoroughly and Systematically from All Study Participants

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

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

4.1.3 Use Patient-Reported Outcomes When Patients or People at Risk of a Condition Are the Best Source of Information

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

4.1.4 Develop and Implement a Dissemination Assessment to Achieve Broad Awareness of Study Results

PCOR research proposals and protocols must include an assessment that describes how the project and the composition of the research team supports dissemination and the anticipated facilitators, barriers and potential strategies for dissemination to key stakeholder groups, including patients and individuals at risk of a condition, clinicians and other healthcare system staff, and policy leaders. Effective dissemination includes the reporting of results in a manner understandable to each target audience, information regarding the relevance of the results for decision-making (recognizing that research findings from a single study alone should not necessarily affect decision-making or practice), along with attention to how the results can be incorporated into health decision-making if applicable. The plan must specify how the dissemination strategy is expected to affect the identified health decisions and how dissemination engages the study participants or the population of interest. Requiring research dissemination, as well as engagement of patients and other stakeholders at this stage of research, represents a cultural shift for many institutions and researchers.

Rationale for the Standards

PCORI must ensure that research proposals and protocols clearly identify the relevant patient populations and those of their health decisions that will be affected by the research.

A focus on patient-centered outcomes is a defining characteristic of PCORI, one that sets it apart from other research funding organizations. Inclusion of patient-centered outcomes is therefore a necessary component of PCORI-funded research. (See *Enhancing the Patients Voice and The Design and Selection of Patient-Reported Outcomes Measures* http://www.pcori.org/what-we-do/methodology/.) The outcomes that patients value directly are distinct from the laboratory tests, examination findings, and other surrogate, substitute, or physiologic outcomes that clinicians may consider important. A patient-centered outcome must meet the following test: "Were it to be the only thing that changed, patients would be willing to undergo a treatment with associated risk, cost, or inconvenience. This would be true of treatments that ameliorated symptoms or prevented morbidity or mortality. It would not be true of treatments that lowered blood pressure, improved cardiac output, improved bone density, or the like, without improving the quality or increasing the length of life."(20, 21)

A Story

"Hi, Sandy, what are you working on these days?"

Sandy often stopped to chat with her older neighbor Judy as she took her dogs out for a walk. She'd been hoping she might see her today. She was just wrapping up a grant proposal for a study on heart failure. Since Judy was being treated for heart failure, Sandy figured she'd be interested.

"We've got our fingers crossed, hoping for funding of a new heart failure study."

"Oh? What's it about?"

"We're looking at the clinic's updated patient management program. We'll be tracking hospital readmissions, ejection fraction, clinic visits, and other information to see what seems to work best."

"That's nice. Will you know if it can help me feel like cooking dinner or have the energy to walk to the bus stop?"

"Um, no," Sandy paused. "Those aren't fields in our electronic medical records."

Many (though not all) meaningful and important patient-centered outcomes, such as symptoms, are best reported by patients themselves. Pain and some other outcomes cannot reliably or accurately be assessed by any means other than direct patient report. If informants from the study population identify outcomes that can only be ascertained by self-report, then inclusion of patient-reported outcomes is essential to patient-centeredness. Even when other sources can provide meaningful outcomes data, patient reports represent the patient perspective and so add value.

Researchers must also incorporate the patient perspective when selecting the type and timing of interventions or exposures to test and the comparisons to make in a study. The selection of appropriate research designs is discussed in more detail in Chapter 6, and specific research method issues are presented in Chapter 7.

To complete the research continuum from the patient-centeredness perspective, an *a priori* dissemination strategy for integrating study results with related work is necessary to optimally affect identified health decisions and outcomes.

While this is the first step to achieving patient-centeredness and relevance in a study, there is a paucity of evidence regarding best practices for patient engagement.(22) The proposed standards in this area thus require investigators to describe in detail in their research proposals and in their publications the ways in which they approached engagement, while avoiding prescribing a particular approach. It is expected that this will focus attention on the importance of this engagement and facilitate evaluation, thus helping to create evidence regarding the effectiveness of different approaches.

To be patient-centered, PCORI and researchers must also engage "informants" in the design, conduct, and dissemination phases of research. Informants are people who are representative of a specific population but are not necessarily study participants. For some populations—for example, children or cognitively impaired persons—informants also include surrogates and caregivers. Although not the focus of this standard, informants representative of other stakeholder groups may also be engaged in the research process.

Further, researchers must ensure that study participants are representative of the spectrum of the population facing the health decision of interest. Work performed for PCORI as background for this standard has identified specific strategies for involving people who have been historically underrepresented in research or who are considered to be hard to reach (*Integrating Patients' Voices* http://pcori.org/assets/pdfs/Integrating%20Patients%20Voices.pdf). Although there are a number of guideline and recommendation documents regarding patient engagement in research,(23, 24) the quality and quantity of empirical evidence regarding patient engagement in research are insufficient to make a full assessment.(24, 25) The first set of patient-centeredness standards for PCOR thus directs researchers to rigorously formulate and describe their methods of patient engagement,

without prescribing a specific approach. Detailed and consistent descriptions of the strategies employed in funded studies will facilitate efforts to evaluate the effectiveness of alternative strategies for engaging patients and other stakeholders. (See *Integrating Patients' Voices* and *Eliciting Patient Perspective* http://www.pcori.org/what-we-do/methodology/.) Appendix C presents a brief narrative review of the Committee's current approach to patient engagement and patient-centered outcomes.

Research Gaps and Future Work

Evidence in a number of areas relevant to the proposed standards is limited, and the standards will evolve over time as additional information becomes available. The Committee intends to focus on three important gaps in knowledge:

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

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

While this set of standards focused on patients (and their surrogates and caregivers), research is needed to clarify which other stakeholder groups should be engaged in which research activities and how to do so. We also need to learn how to balance and reconcile the inputs from the various stakeholder groups in the design, conduct, and dissemination of PCOR.

Recommended Actions

The following actions are recommended for PCORI to adopt in order to better enable patient engagement in its activities:

- Support training in patient engagement methods for investigators and patient informants.
- Improve the patient-reported outcomes (PRO) evidence base by supporting research on methods for assessing measurement properties (based on qualitative and quantitative evaluations), score interpretability, meaningfulness of score changes, and strategies for minimizing and interpreting missing PRO data in PCOR.
- Evaluate patient dissemination activities, and require incorporation in future research of relevant learnings from this evaluation.

Research Recommendations

- Create an infrastructure to support research on patient engagement. To facilitate this, PCORI should:
- Develop a standardized nomenclature for patient engagement methods.
- Develop a sample patient engagement plan to demonstrate the key elements required for patient engagement in the research process. The sample plan should illustrate engagement of both patient informants and study participants.
- Systematically collect information about patient engagement methods from PCORI-sponsored studies.
- Evaluate the effectiveness of patient informant engagement.
- Synthesize results across studies.
- Disseminate findings to improve patient engagement in PCOR.

Chapter 5. Methods for Prioritizing Patient-Centered Outcomes Research

Establishing a specific research agenda is a core duty of PCORI. Unless the mismatch between current research priorities and the information needs of patients and clinicians is addressed, our methodological standards will have little effect. PCORI research must be directed toward providing the answers patients need in order to make health decisions.

The PCORI Board of Governors is charged with developing, refining, informing priorities for, and selecting among research investments. In this chapter, the Committee provides guidance to the Board regarding how the methodological tools for research prioritization might be of use in these processes. The Committee provides a framework and narrative overview of topics related to research prioritization and proposes three standards with supporting recommendations. Elements of this framework have already been used by the Agency for Healthcare Research and Quality (AHRQ) and the Institute of Medicine to identify gaps in the evidence and topics that PCORI should consider.(14, 26) PCORI should work closely with the Methodology Committee to build on previous work and implement the framework efficiently.

Background

The 2010 Patient Protection and Affordable Care Act (PPACA) lists factors that should be taken into account in identifying priorities (Box 5.1) (Appendix E-3). In addition, research funded by PCORI should seek to take into account differences across individuals and groups of individuals (Appendix E-4).

Box 5.1. Research Prioritization Factors

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

While this list includes the effect of the proposed research on national healthcare expenditures as a consideration, elsewhere the legislation restricts consideration of cost. The Committee's view is that in the context of PCOR, cost, like a number of other social barriers, can be a factor in the effectiveness of a given treatment if it influences choices made by patients and clinicians.

"The Institute shall establish and update a research project agenda for research to address the priorities identified under subparagraph (A), taking into consideration the types of research that might address each priority and the relative value (determined based on the cost of conducting research compared to the potential usefulness of the information produced by research) associated with the different types of research, and such other factors as the Institute determines appropriate." (Appendix E-3)

The PPACA makes clear that PCORI must spend its resources effectively and efficiently (see Box 5.1). When there is more than one acceptable research approach available, the potential added cost of the methods should be balanced against the potential value, including the timeliness, of the research results they are likely to produce. Techniques such as Value of Information (VOI) analysis—a statistical method for estimating the average improvement in outcomes that may be expected by obtaining additional information(27)—may be useful in clarifying tradeoffs between study cost and the degree of certainty that is expected from study results.

Framework for Establishing Research Priorities

The focus of the framework is to inform the selection of PCOR research proposals for funding that attempt to answer specific diagnostic, therapeutic, or health system questions. There are four key components to this framework:

1. **Topic Generation.** The purpose is to ensure that a sufficient number and range of topics are considered before topics for research funding are selected. Engagement of multiple stakeholders, especially patients, is critical at this stage. (See *Methods for Involving Patients in Topic Generation* http://www.pcori.org/what-we-do/methodology/.)

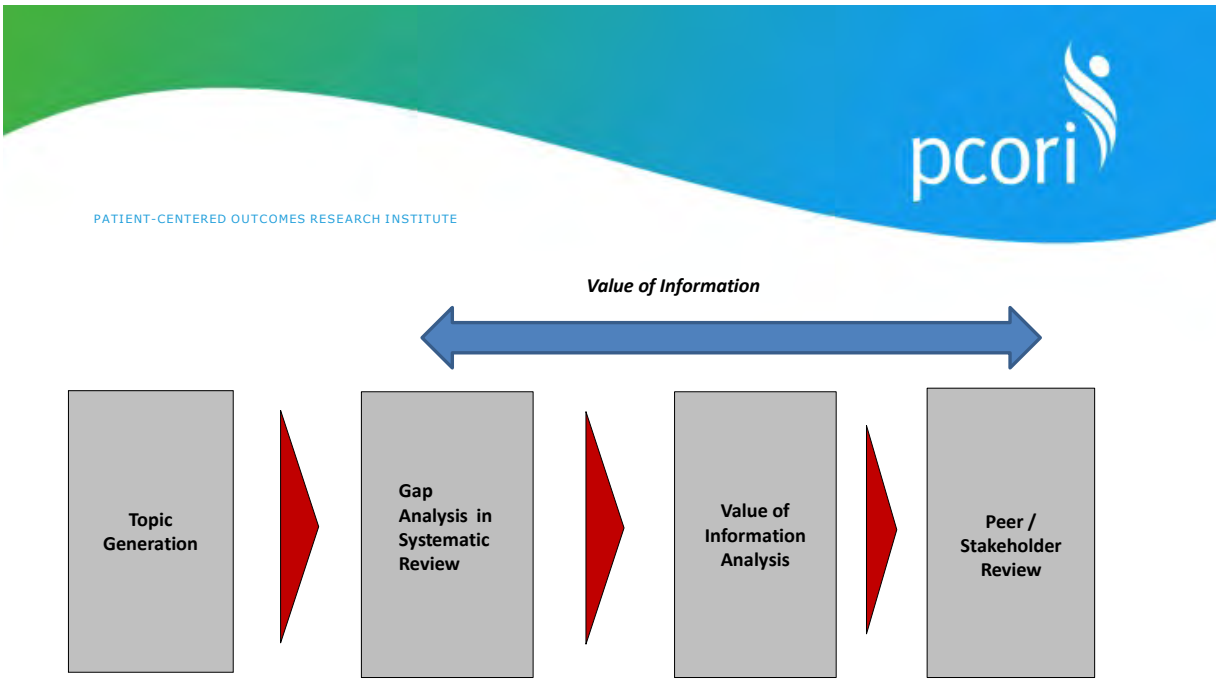
2. **Systematic Review and Gap Analysis.** By involving patients and other stakeholders in developing questions for a systematic review, researchers can compare what people want and need to know with what is and is not known. As a result, systematic reviews can identify gaps in knowledge that underlie uncertainty among patients and clinicians. Sometimes, systematic reviews can generate new questions. For example, a pooled analysis of several

studies can reveal an important finding that was not evident in the individual studies. (See *Prioritizing Future Research* http://www.pcori.org/what-we-do/methodology/.)

3. **Value of Information Analysis.** Value of information analysis may be used to identify questions that have the greatest potential to improve population health by considering uncertainty in the health benefits and risks associated with alternative treatment choices, the ability of research findings to alter that uncertainty, and the resulting care decisions. (See *Value of Information and Research Prioritization* and *Value-of-Information Analysis* http://www.pcori.org/what-we-do/methodology/.)

4. **Peer and Stakeholder Review.** Peer and stakeholder review is the final stage in selecting research proposals for funding. (See *Peer Review – A Research Priority* http://www.pcori.org/what-we-do/methodology/.)

Figure 5.1: Framework for Research Prioritization



The Committee adopted standards for only two of the four components listed above and acknowledges that the evidence supporting these specific research prioritization methods in the standards is incomplete. Therefore, the Committee proposes that PCORI support empirical research to assess and improve research prioritization methods (see Recommendations for Research).

Standards for Prioritizing Research

5.1.1 **Use Systematic Reviews to Identify Gaps in Evidence**

Gap analysis of systematic reviews should be used as part of the process of identifying and prioritizing research gaps to establish funding priorities by PCORI.

5.1.2 **Protect Independence in Peer Review of Research Funding Proposals**

Adopted methods of peer review should aim to safeguard independence between reviewers and those being reviewed.

5.1.3 **Ensure Adequate Representation of Minorities and Disadvantaged Segments of the Population in Peer Review of Research Funding Proposals**

Approaches to topic generation in PCOR should involve both consultative and collaborative functions.

Topic Generation

While the Committee proposes neither standards nor recommended actions at this time regarding topic generation, it believes that the subject requires additional research. Including patients in topic generation is unconventional. Topic selection is usually done by researchers or sponsors, and while they may believe they know what patients want, their choices may be influenced by personal or commercial interests. Without adequate input from patients, research priorities may not fully reflect patient perspectives on potential benefits or risks, ultimately impeding the uptake of research discoveries.

Some empirical research, mostly conducted outside the United States, has shown that patient involvement can produce more relevant research questions and more pertinent data outputs.(28, 29) The Committee commissioned a review that examined some of these approaches in detail (*Methods for Involving Patients in Topic Generation* http://www.pcori.org/what-we-do/methodology/). The Committee believes that PCORI should test and develop existing alternative approaches and novel

methods to obtaining patient input in research topic generation in order to determine which approaches work best for specific patient populations (see Recommended Actions).

Systematic Review and Gap Analysis

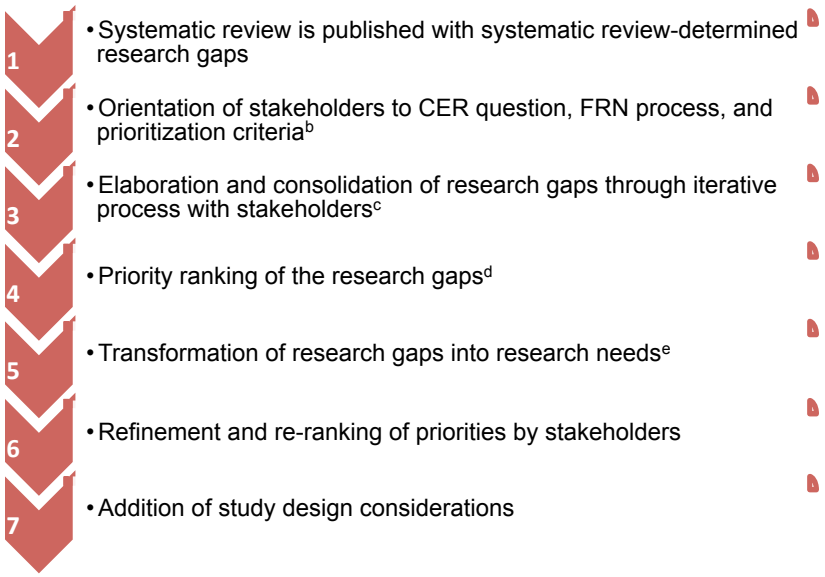
Gap analysis of systematic reviews should be used as part of the process of identifying and prioritizing research gaps to establish funding priorities by PCORI.

Rationale for the Standard

Using a systematic approach to identify gaps in the existing literature and deficiencies in completed studies should reduce investments in research that are unlikely to help answer important questions, while avoiding potentially unethical patient recruitment into unneeded studies and fostering transparency and accountability in funding prioritization.

While the Committee does not formally endorse a specific approach, it notes that AHRQ has advanced the field in the United States significantly with its Future Research Needs (FRN) Approach. In 2009 AHRQ selected eight Evidence-based Practice Centers (EPCs) to conduct comparative effectiveness reviews and complementary FRN (Figure 5.2).(30) Other national and international organizations have used similar approaches.

Figure 5.2. Adapted Example of a Future Research Needs Process[a]



[a] This figure was adapted from a draft AHRQ FRN methods paper.(31)
[b] May include identification of additional research gaps.
[c] Reduction through topic consolidation, preliminary prioritization, and consideration of ongoing research (duplication criteria).
[d] Research gaps that address specific methods issues would not use PICOTS framework.
[e] May require iterative steps.

Value of Information Analysis

The Committee proposes no standards for but recommends two actions regarding value of information (VOI) analysis. VOI analysis takes into account the research prioritization factors listed in the PPACA by integrating them into a single measure – the expected (average) increase in population health that might be expected from a research project. VOI analysis is the idea, rooted in statistical decision and economic theory, that the value of information can be defined by the average improvement in outcomes expected by obtaining additional information.(32, 33) Background papers commissioned by the Committee for this report provide an overview of this approach and highlight specific issues that must be addressed in its potential application by PCORI.

Peer Review

The Committee proposes two standards regarding peer review. Despite its central role in scientific discourse and decision-making, peer review has had little attention as the subject of research. Rigorous experiments testing alternative approaches to peer review are rare; most peer-review practices are maintained by convention.

PCORI has particular advantages and responsibilities in developing its approach to peer review. For example, incorporating patient stakeholders presents both a new opportunity and a challenge for peer review. A paper commissioned by the Committee presents the limitations of current methods of peer review for research proposals and the ways patients or their advocates have been involved. (*Peer Review – A Research Priority* http://www.pcori.org/what-we-do/methodology/) Review practices vary substantially, and in the absence of evidence it is not possible to recommend one mode over another, or even to recommend when peer review of proposals is the best possible way to allocate funding and other resources. Nevertheless, effective peer review is indispensable to the conduct of PCOR, and independence between those being reviewed and those reviewing proposals must be safeguarded.

Rationale for the Standards

Lack of independence is likely to create circles of insiders, propagate inbreeding of ideas, and decrease scientific rigor and research quality. Most research funding agencies try to exclude scientists who have perceived conflicts of interest from reviewing applications, although definitions of conflicts vary. In a densely connected world of scientific research, links between investigators are often difficult to identify and regulate, and there is variability in how investigators with potential conflicts of interest are excluded from the process.

Including patients or their representatives in the peer review process will introduce new conflicts of interest due to potential links of these stakeholders to particular investigators, advocacy groups, and other patients with the condition. Methodological research is needed to understand the potential threats to independence that could arise from conflicts of interest when patients and related stakeholders are involved in the peer review process.

In addition, the adequate and appropriate representation of minorities is routinely considered as a standard in NIH funding opportunities and many other funding agencies. Differences in disease burden may suggest different research priorities. There is empirical evidence that minorities and disadvantaged segments of the population are underrepresented in many areas of clinical research. Whether or not the peer review process is responsible, it may help remedy the imbalance.

In conclusion, the dearth of strong scientific evidence on the most effective methods for peer review with respect to independence and representativeness of reviewers and the efficiency of the

process, along with special considerations relating to minority populations, points to the need for additional research.

Recommended Actions

To begin a program of patient-centered research prioritization, PCORI should take a number of actions, including the following.

- Work closely with the Methodology Committee to build on previous work and implement the framework efficiently.
- Base all PCORI targeted funding opportunity announcements on evidence gap analysis.
- Require that applicants demonstrate how their proposed research fills a research gap for non-targeted funding opportunity announcements.
- Support education and training activities to broaden the base of individuals prepared to apply and evaluate VOI.
- Maintain peer review processes that avoid interference of participants and stakeholders with potential conflicts of interest. Peer review should incorporate patient perspectives.

Research Recommendations

- Encourage intra- and extramural research in the development and practical application of VOI methods for PCOR, including through studies that examine the contribution of VOI methods to research prioritization when used in conjunction with other approaches to research prioritization.
- Support empirical research to assess and improve research prioritization methods for use by PCORI.
- Support extra- and/or intramural research to establish a best practice approach to consultative and collaborative patient engagement in topic generation that is suitable for the heterogeneity of the US patient population.
- Study the employment of research gap analysis to continue to develop the empirical evidence on its use.
- Encourage studies, ideally with experimental designs, that assess different methods for engaging patients with diverse views and preferences and funneling their input into the peer review process in a consultative manner.

Chapter 6. Choosing Data Sources, Research Design, and Analysis Plan: Translation Framework and Development of a Translation Table

The choice of a study design is one of the most critical factors in PCOR. The legislation expected the Methodology Committee to address this question, calling on it to develop "*a translation table that is designed to provide guidance and act as a reference for the Board to determine research methods that are most likely to address each specific comparative clinical effectiveness research question*" (Appendix E-5). The choice of study designs has practical implications for the timeliness, validity, and relevance of PCORI's specific research project agenda.

The translation table serves to guide the design of new comparative clinical effectiveness research studies for specific research questions. Clearer guidance regarding the selection of appropriate research designs for specific research questions could avoid studies that are inappropriate for the research question and could direct researchers to designs that reach the "right" answers sooner, thus improving the efficiency of research. When research designs clearly match the questions patients and their healthcare advisers consider important, research results should be more readily accepted and implemented. A translation table could help build a comprehensive research program recognizing and balancing the inherent tradeoffs of each study design and analytical methodology.

Research funding agencies, including PCORI, could require that a translation table be used for developing research proposals and for evaluating the quality of such proposals. The legislation names the PCORI Board as the primary audience for the translation table, but other entities that generate or interpret comparative effectiveness research/patient-centered outcomes research are likely to be guided by such an instrument. These include applied researchers, regulatory and funding agencies, and payer and provider organizations.

Very few published articles mention the concept of a "translation table," and stakeholders have varying opinions about what it should include (see Appendices 6-1, 6-2, and 6-4). The lack of expert consensus required the Committee to pursue its own approach to developing a translation table, using information from the public to inform our decisions about the scope and form a translation table could have. We developed a set of principles and a *translation framework,* which we refined by applying case studies contributed by Committee members and members of the public (see Appendices 6-3, 6-5).

Scope

The Methodology Committee decided that the translation table should address two main tasks:

1. Choosing a basic study design. By basic design, we mean whether the study is experimental or observational. Other study designs, such as systematic reviews and decision models, are also within the scope of PCORI but are not addressed in this version of the translation table.

2. Determining additional design details. Once a decision is made to conduct an observational or experimental study, a number of options about study design need to be considered and weighed. In the last section of this chapter, we provide a detailed example of choosing additional design features for a randomized trial.

Table 6.1 shows two basic design types—the randomized clinical trial design and the nonrandomized study—used to assess the effectiveness of therapies. Within each basic design, three examples of more specific designs are described. A translation table would be used to decide first between a randomized or nonrandomized design; once that decision was made, it could then be used to choose additional, more specific design features.

Table 6.1. Selected Study Designs for Assessing the Effectiveness of Therapeutic Interventions

Study Design	Description
Randomized clinical trial	A study in which *randomization* is used to assign study subjects to interventions and which can range in the amount of control exercised by the investigators. Randomized controlled trials determine whether a clinical intervention is effective under optimal circumstances. This is achieved by using rigorous inclusion and exclusion criteria and a highly controlled research environment. Pragmatic randomized studies determine the harms, benefits, and costs of an intervention as it would occur in routine clinical practice. This is achieved by including a combination of a broader range of subjects, study sites, and outcomes.
Cluster randomized clinical trial	A randomized clinical trial that groups subjects according to "clusters," such as clinic site or community. *Randomization* is used to assign entire clusters to interventions. This design is useful when evaluating health policies or when randomization at the subject level is not possible.
Crossover trial	A trial in which the subject acts as his/her own control. This is accomplished by having each subject receive a sequence of interventions over periods of time, crossing over to an alternative intervention as part of the sequence. The order of intervention is assigned *randomly*. The N-of-1 trial is a special case of a crossover trial that compares two or more interventions for a *single* patient.
Delayed start trial	A form of crossover trial conducted in two phases, sometimes used to evaluate whether an intervention acts by reducing symptoms (Phase I) or by modifying disease (Phase II). In Phase I, subjects are *randomized* to receive either the intervention or comparator; in Phase II, all subjects receive the intervention.
Nonrandomized study	A study that does not use randomization to assign participants to intervention arms.
Cohort study	A nonrandomized study of subjects with a common feature or condition. Subjects are *observed* to receive specific interventions. Data may be collected and evaluated prospectively or retrospectively. Efficient sampling designs are available for cohort studies, including case-control studies, case-cohort studies, and 2-stage sampling designs.
Self-controlled designs	A nonrandomized study in which subjects act as their own controls by observing a sequence of treatments over periods of time crossing over to an alternative interventions as part of the sequence.
Controlled time-trend analysis design	A nonrandomized study in which sudden changes or shocks in the healthcare settings are exploited to compare outcomes before the change with outcomes observed after the change. When subjects who have not been exposed to the shock over the same time period are available, this design is referred to as a difference-in-difference or quasi-experimental design.

Emerging Principles

The Committee recommends the following principles in the translation framework.

1. Keep the research question and the methodology separate: In order for the translation table to function in the way envisioned by the legislation, the Committee believes that it is important to separate the development of the research questions from the task of finding the best methodology to answer the question.

The Committee views the research methodology as the means to answer a research question as well as possible, not as a factor that should influence the choice of research question. Problems occur when the choice of study design or research question is driven by data availability. Defining the question should not be limited by concerns about eventual methodological constraints. In PCOR, identifying decisions and defining a patient-centered research question should come first.

2. Focus on clarifying tradeoffs: After a research question is defined, choices have to be made about the type and level of evidence needed to inform the decisions that it was intended to address. This will direct the choice of research design and analytic strategy. Preference should be declared on a series of factors including (in no specific order) the timeliness of findings, their representativeness, their validity, the ability to identify subgroup effects, and others. Such *study characteristics* (see Box 6.1) need to be defined with stakeholder input as they will substantially influence the utility of the results for decision-making. Clearly articulating the tradeoffs between these choices will bolster the transparency of the analytical approach.

Box 6.1. Examples of Study Characteristics

Intrinsic study characteristics

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

Extrinsic study characteristics

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

3. Expect that several versions of a translation table might be needed for different decision-makers and different *research categories* (see Box 6.2): Different decision-makers may have different evidentiary needs. The play of study characteristics as well as the preferences of decision-makers may depend on whether the topic is about a choice of diagnostic approaches, an attempt to estimate risk or predict future complications, a choice of surgical versus nonsurgical treatment, or a choice of ways to change the way care is delivered. Eventually each kind of researchable question will require its own translation table that can provide meaningful advice on its specific methodological requirements.

4. Place individual research studies in the context of a research program: The legislation asked for a translation table that helps match the best methodology to a specific research question. This implies a one-to-one relationship. It is widely accepted that most research questions can be answered in several ways. A research program may, for example, include an effectiveness study based on secondary healthcare data, a detailed interview study, and a randomized trial to balance population representativeness, timeliness, depth, and validity for informed decision-making. The translation table should help identify the tradeoffs between different approaches to answering a question. For example, suppose a new surgical procedure to repair heart valves that is less invasive than the standard surgery has been developed but requires specialized surgical training and skill and the participation of cardiac surgery team. A randomized trial may be required to establish the benefits or harms of the new procedure compared to the standard procedure under these ideal conditions. Regulators are likely to be very interested in the outcome of this type of design. An observational study may also be needed to determine the safety and effectiveness of the new procedure compared to the standard approaches when the procedure becomes more widely available. Patients and professional societies may be interested in the outcome of this study—patients to determine if the new procedure is a good choice for "people like me" and professional societies to determine if and how best to adopt the new procedure.

5. The choice of study design must take into account the state of the art of research methodology: Over the past 20 years, the choice of study design has been debated intensely in scientific and, more recently, political circles. These discussions often reiterate commonly held beliefs about randomized controlled trials (RCTs) and observational studies. One asserts that RCTs are less relevant to decision-makers than observational studies. Another is that

observational studies nearly always suffer from serious flaws that render them invalid and even irrelevant. In fact, many RCTs have proven to have long-lasting value in clinical decision-making. In many fields, critical evidence comes from RCTs, many of them conducted in patient populations and circumstances that are broadly applicable. Observational studies have been extremely valuable as a complement to RCTs, helping to determine under what circumstances and to what patients the findings of RCTs are applicable. Serious errors in clinical practice can be due to overreliance on narrowly focused RCTs or on flawed observational studies, but widely cited examples do not represent the potential of these basic designs to contribute to PCOR.

The familiar overgeneralizations do not take into account developments in research methodology that can make randomized studies more relevant, timely, and flexible, and can improve the validity of observational studies. In particular, the use of observational studies to make causal inference is potentially much stronger than it has been in the past. Many of the standards that we developed address ways to improve the value of observational studies as a substitute for RCTs for questions about comparative clinical effectiveness. Decisions about study design need to take into account these standards, described in Chapters 7 and 8, below, and the advances in methodology they reflect.

Translation Framework

A single translation table would not work for all types of research questions. We developed a conceptual model of the translation task based on three concepts, defined below (Box 6.2).

Box 6.2. Terms for Describing the Translation Table

Research Category: Breaking research into categories will make each translation table more informative and less complex. For example, original research on the effectiveness of therapeutics will require a different approach than will research on the effectiveness of diagnostic tests or imaging. Research categories and their relationship to each other are not yet defined and will develop as the translation framework develops.

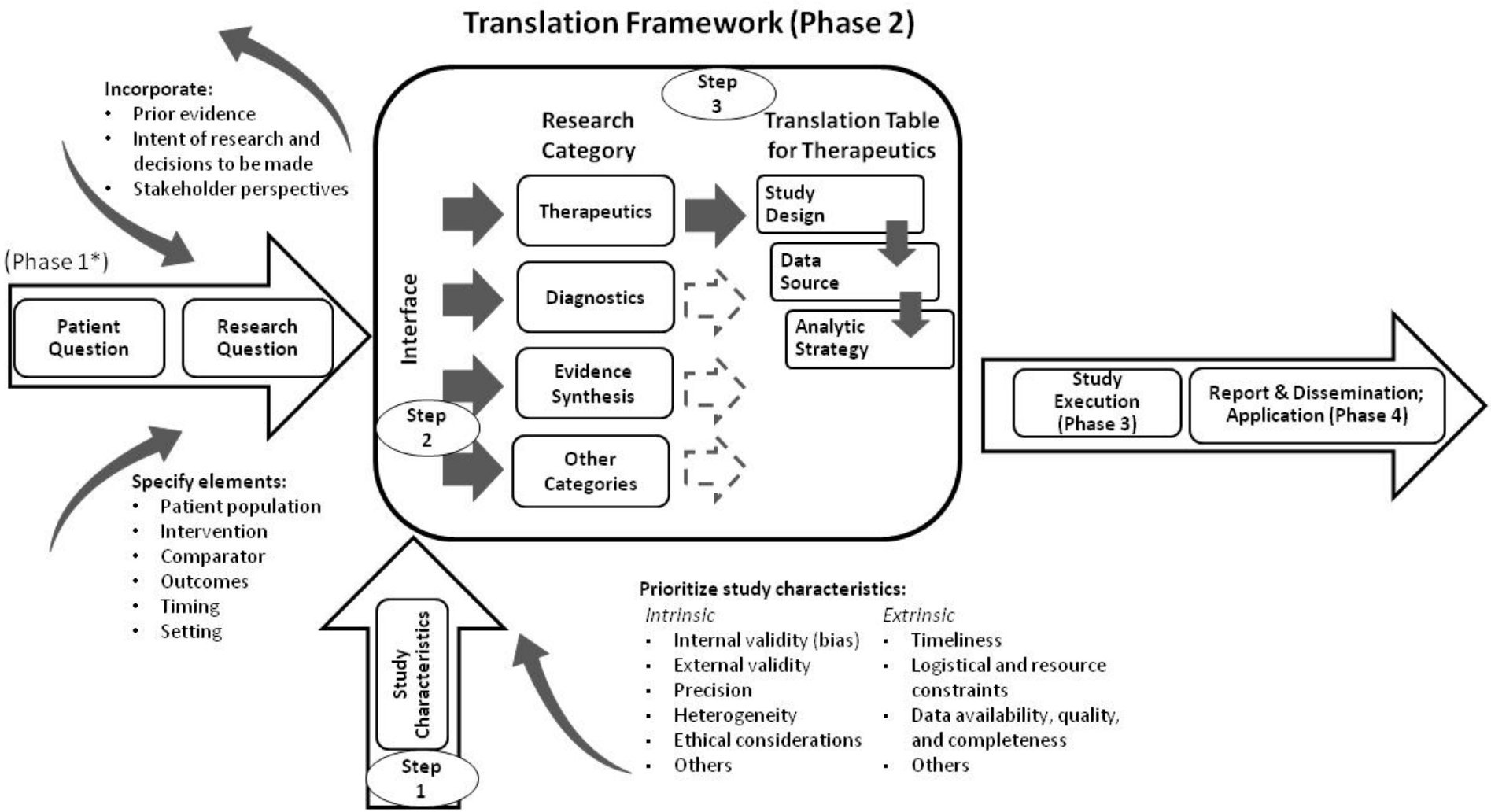
Translation Framework: The translation framework provides the theoretical underpinning and organizing structure for the translation table that will help PCORI identify the range of appropriate research designs and analytic approaches to answer specific patient-centered research questions. It includes a set of study characteristics that can be used to guide the user in making choices in study design and analytic methods based on current scientific knowledge, research categories, and the corresponding translation tables.

Translation Tool: Although the legislative mandate is to create a translation table, the usefulness of a literal "table" format may be limited. As the combination of study characteristics, research categories, and translation tables may be complex, a translation tool, rather than a table, would make the translation framework more accessible to users.

The Committee envisions the translation tool as a dynamic implementation of the translation framework to help users apply it to specific research questions. Such a tool might take the form of weighting or optimization algorithms that are made accessible to users in electronic formats. The tool would point to one or more recommended designs and to other designs that not acceptable.

The basic structure and function of the proposed translation framework (Figure 6.1) can be summarized in four phases, corresponding with the phases described in Chapter 3, Table 3.1.

Figure 6.1. Translation Framework (Phase 2)



* See Table 3.1, Phases of PCOR

The translation framework begins with the patient's healthcare decision and assumes that a patient-centered research question has been precisely specified which includes critical appraisal of prior studies; an assessment of what is known, what is unknown, and why; and consideration of the decision the study is meant to inform (see the first three activities in Table 3.1 Phase 1). Multiple perspectives may shape the research question, including those of patients, clinicians, researchers, policy-makers, and others.

It is part of the premise of PCORI that a lot of well-designed research is not designed to inform and does not inform patients' decisions. Formulating decisions and defining questions for research are complex processes requiring expertise and open-mindedness,(34, 35) careful formulation of research questions (above, Chapter 3), and a commitment to patient-centeredness (above, Chapter 4).

The focus of the translation framework is **Phase 2.** The framework diagram highlights 3 steps:

Step 1: Once a patient-centered research question has been specified, the next step is to determine the type of evidence that will be required to address the question and meet the needs of stakeholders. The type of evidence that the study is expected to produce can be determined by a set of study characteristics that need to be considered and prioritized to ensure that the study will yield meaningful results. These include intrinsic characteristics (those that can be influenced through study size, selection of study subjects, and research methodology) and extrinsic characteristics (those that are less directly under the control of the researchers). Examples are provided below (see Box 6.2).

Policymakers, clinicians, researchers, and patients may prioritize study characteristics differently, sometimes in incompatible ways. Therefore, prioritizing study characteristics is an important step in patient-centered outcomes research, making transparent the differences in evidentiary needs. This process may ultimately result in a set of studies all on the same question but with slightly different goals.

Step 2: Completion of steps 1 and 2 lead to the interface matching the research question and defined study characteristics to an appropriate research category (e.g., effectiveness of therapeutics, effectiveness of diagnostics, and others).

Step 3: Each research category has an associated translation table. The translation table outlines the tradeoffs among various methodological approaches for a specific setting that is defined by the research question and the relative importance of the study characteristics. The table could indicate several acceptable designs and analytic strategies, perhaps ranked by their appropriateness to the intended study. It is expected that narrowly defined research categories could provide more prescriptive guidance, while broader categories could describe the advantages and disadvantages of several design and analytic options.

Many choices that need to be made in designing an appropriate study are not binary decisions, but rather tradeoffs among the limitations inherent to each design and analysis approach. Every design has characteristics that must be balanced when developing a study to address a particular research question. For example:

- In order for the results of a study to be obtained faster or to maximize external validity, an observational study using secondary data could be considered. However, this design would likely have less internal validity than does an experimental study that uses randomization. The experimental study could fail to address the research question, though, if it is not representative of care outside the controlled research environment.

Example: To determine the effectiveness of a variety of marketed rheumatoid arthritis (RA) therapies and their combinations in various patient subgroups representing the wide range of users in routine care, a large cohort study may provide answers either by utilizing electronic health records or by utilizing RA disease registries. The number of different medication patterns and subpopulations would require a very large "pragmatic" trial (see Table 6.1).(17)

- A study design without a comparator, based on information from a device registry, *might* be acceptable for assessing device failure rates, but not to assess device effectiveness.

Example: A large health system instituted a registry of all patients undergoing joint replacement surgery with subject level information about the patient's condition, type of surgery, device-specific identifiers, and postoperative follow-up. Device failure rates could be tracked over time for specific devices and correlated with patient specific information. However, for determination of whether the device failure rates are too high, comparators are required.

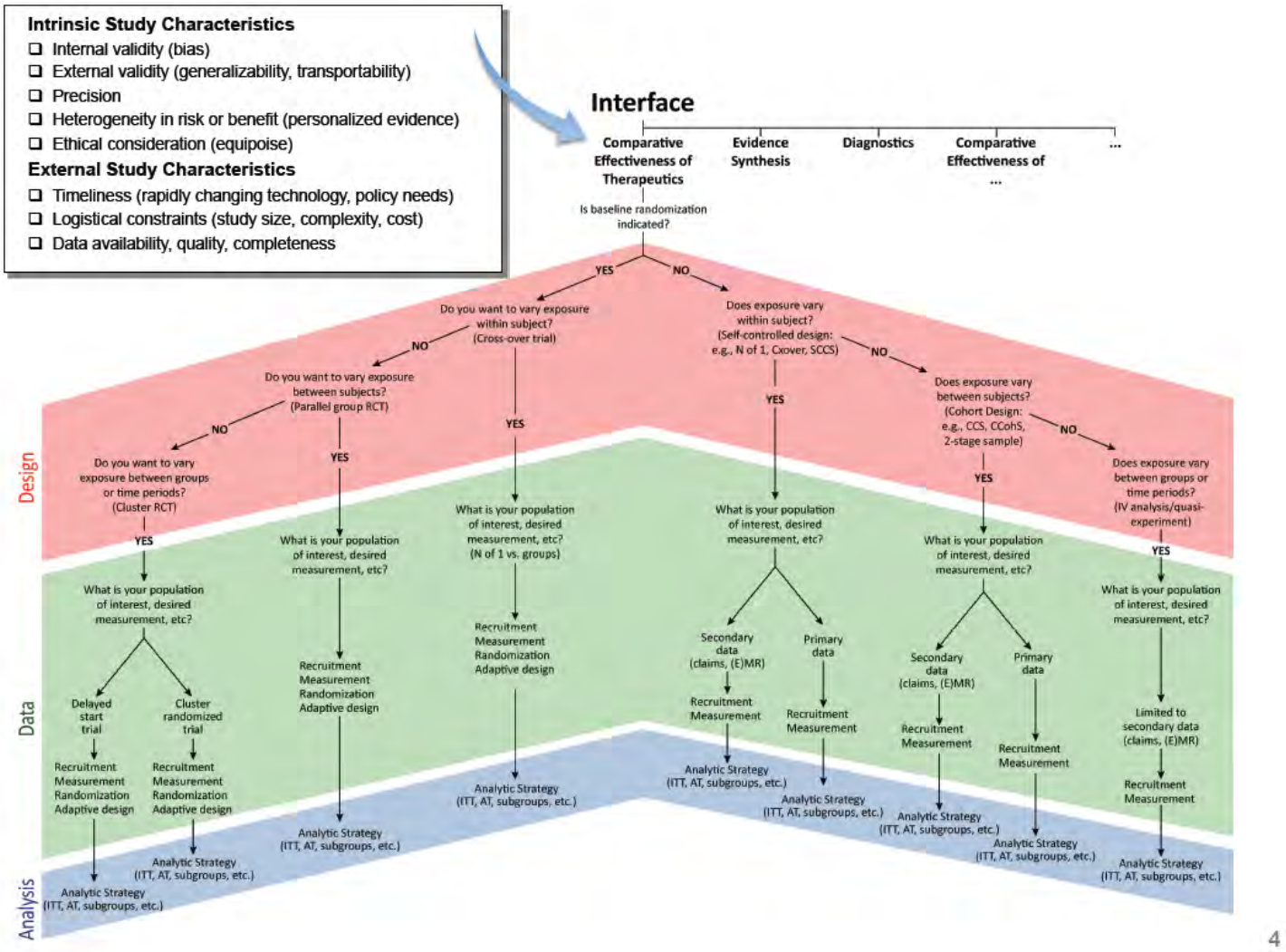
- Often, logistical issues can be more challenging than scientific issues. For example, if only a limited number of patients are available to study a specific condition, an efficient sampling strategy within existing healthcare data sources may be necessary to successfully conduct the study.

Example: Women with specific deleterious genetic mutations, identified as *BRCA1* and *BRCA2,* have increased risks for developing breast and ovarian cancer. However, the prevalence of mutation carriers is low in the population. To determine the comparative effectiveness of three medications to reduce risk for breast cancer in women with deleterious mutations, an investigator could recruit potential participants from a registry or health system database, provided patient privacy is maintained.

Use of the Translation Framework

There are several ways to convert the translation framework into a usable translation tool. For example, a decision tree could be devised to guide the sequential decisions to be made by a researcher for a specific research category. Such a decision tree for the effectiveness of therapeutics is illustrated below.

Figure 6.2. Decision Tree for Comparative Effectiveness of Therapeutics



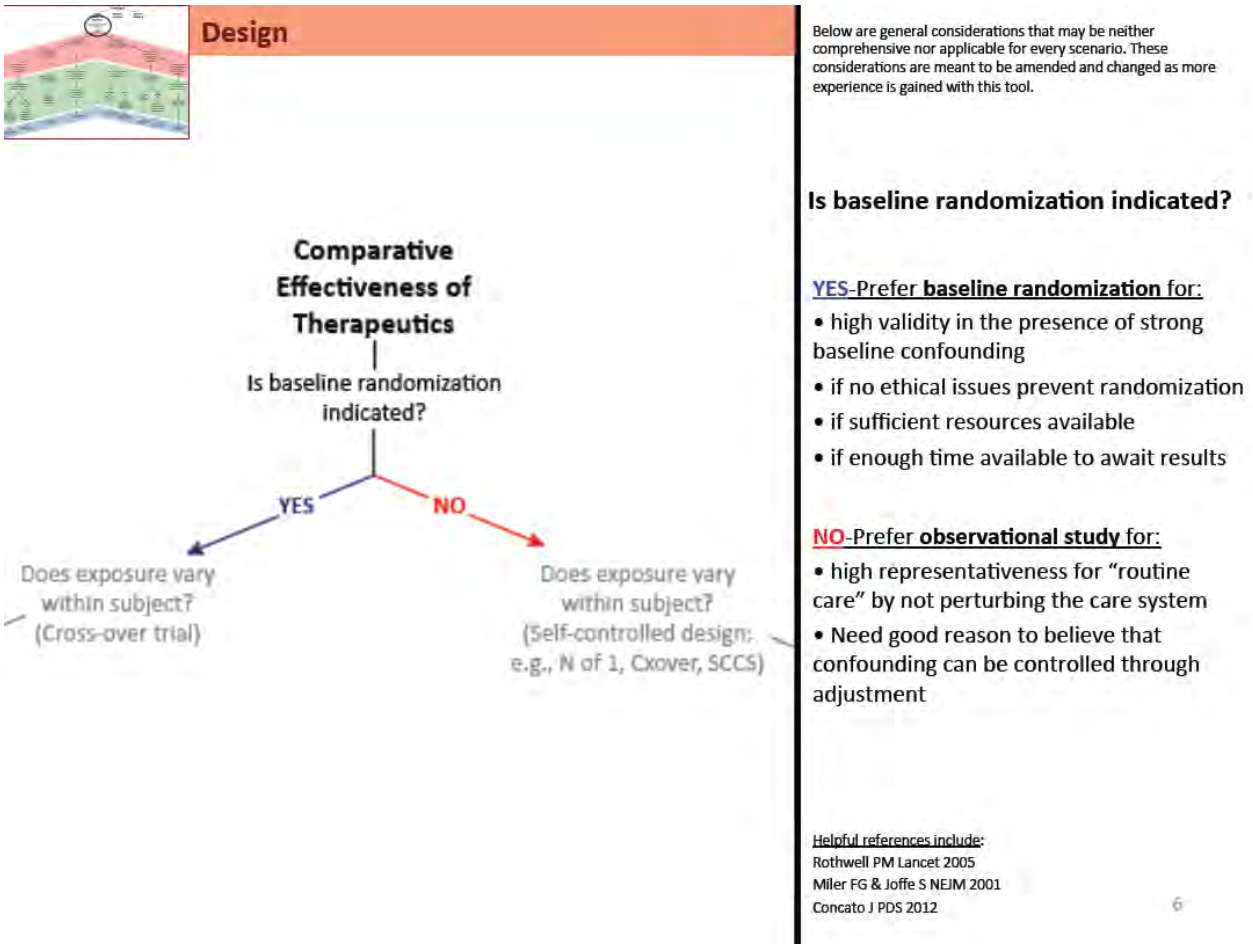
Abbreviations: RCT = randomized controlled trial; AT = as treated analysis; ITT = intention to treat analysis; EMR = electronic medical records; IV = instrumental variable; Cxover = case-crossover study; SCCS = self-controlled case series; CCS = case-control study; CCohS = case-cohort study.

It is assumed that the user of the tool will start at the interface after 1) the research question is precisely formulated and 2) the desired study characteristics are prioritized. At the interface, 3) the research question and defined study characteristics are matched to the most appropriate research methodology, and 4) the research dimension is clarified (here: effectiveness of therapeutics).

Choosing the Basic Study Design

The user will then follow to the first decision point in the translation tool. In this example, the first decision point is: ***Is baseline randomization indicated?*** This question is of such fundamental importance and influences most other design and analytic choices that it needs to be answered before moving on. Figure 6.3 below illustrates this branch point:

Figure 6.3. Is Baseline Randomization Indicated?

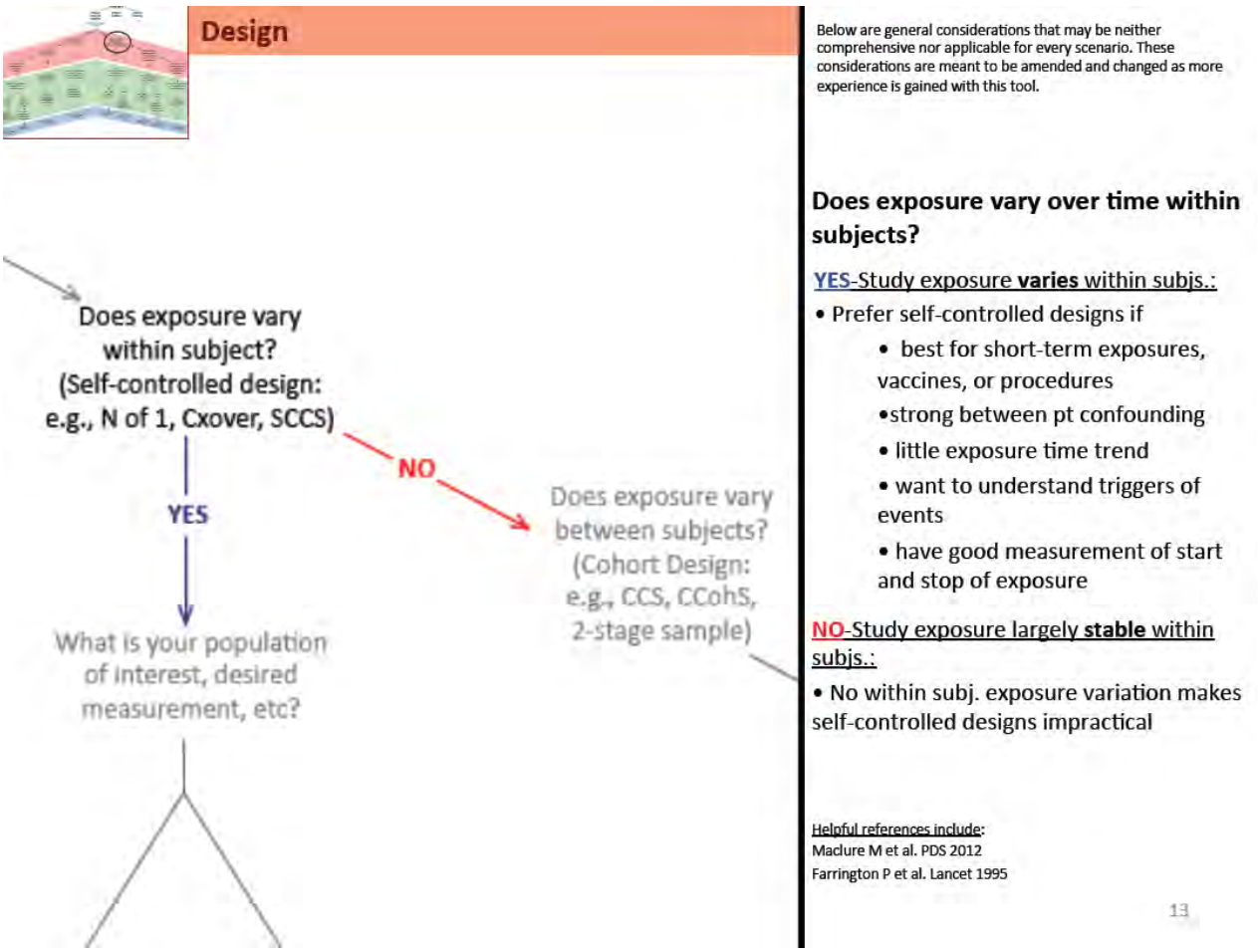


Randomization intends to balance baseline characteristics of research subjects to ensure fair

treatment comparisons. The primary purposes of randomization are considered at this decision point. If an investigator concludes that no observational study will ever be able to balance comparison groups, then the choice for randomization is clear. However there are constraints that may make randomization impossible, including the lack of clinical equipoise and the inability to recruit sufficient numbers of participants.

Determining additional design details. Once a decision is made to conduct an observational or experimental study, a number of options about study design need to be considered and weighed. In this case, the user has decided that it is highly unlikely that confounding can be controlled in a non-randomized study and that generalizability to routine care settings is of high priority.

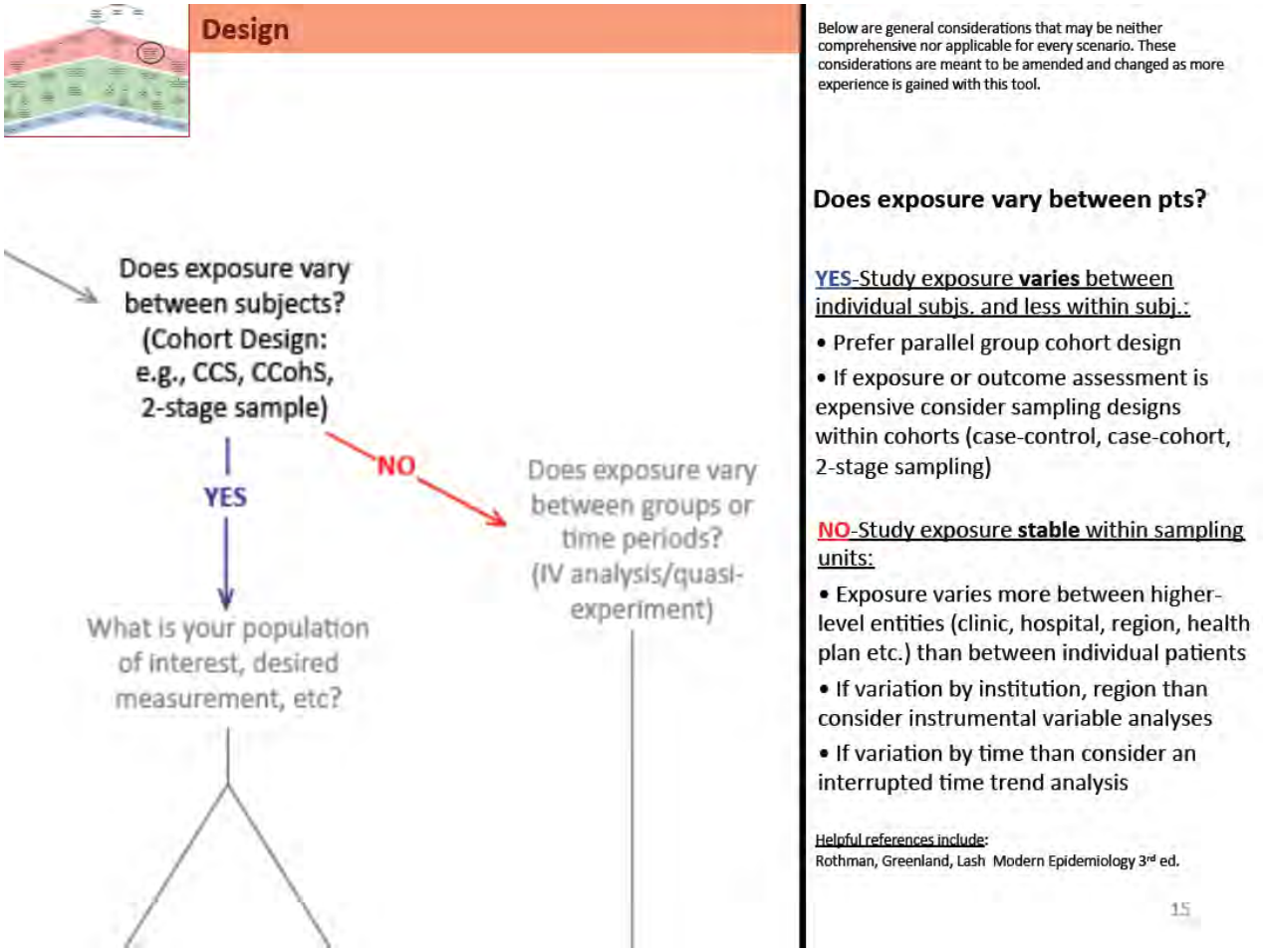
Figure 6.4. Does Exposure Vary Over Time Within Subjects?



The next decision point is: *Does exposure vary within subjects?* Some treatments are used sporadically or just once. If the treatment of interest is short-term or one-time (e.g. antibiotics,

vaccines), then self-controlled designs offer attractive properties, including improved confounding adjustment for time invariant patient factors. Self-controlled designs are in their approach and interpretation of results sufficiently different from other nonrandomized studies that a separate path opens. If an exposure of interest is longer-term, then a cohort study design needs to be considered (Figure 6.5).

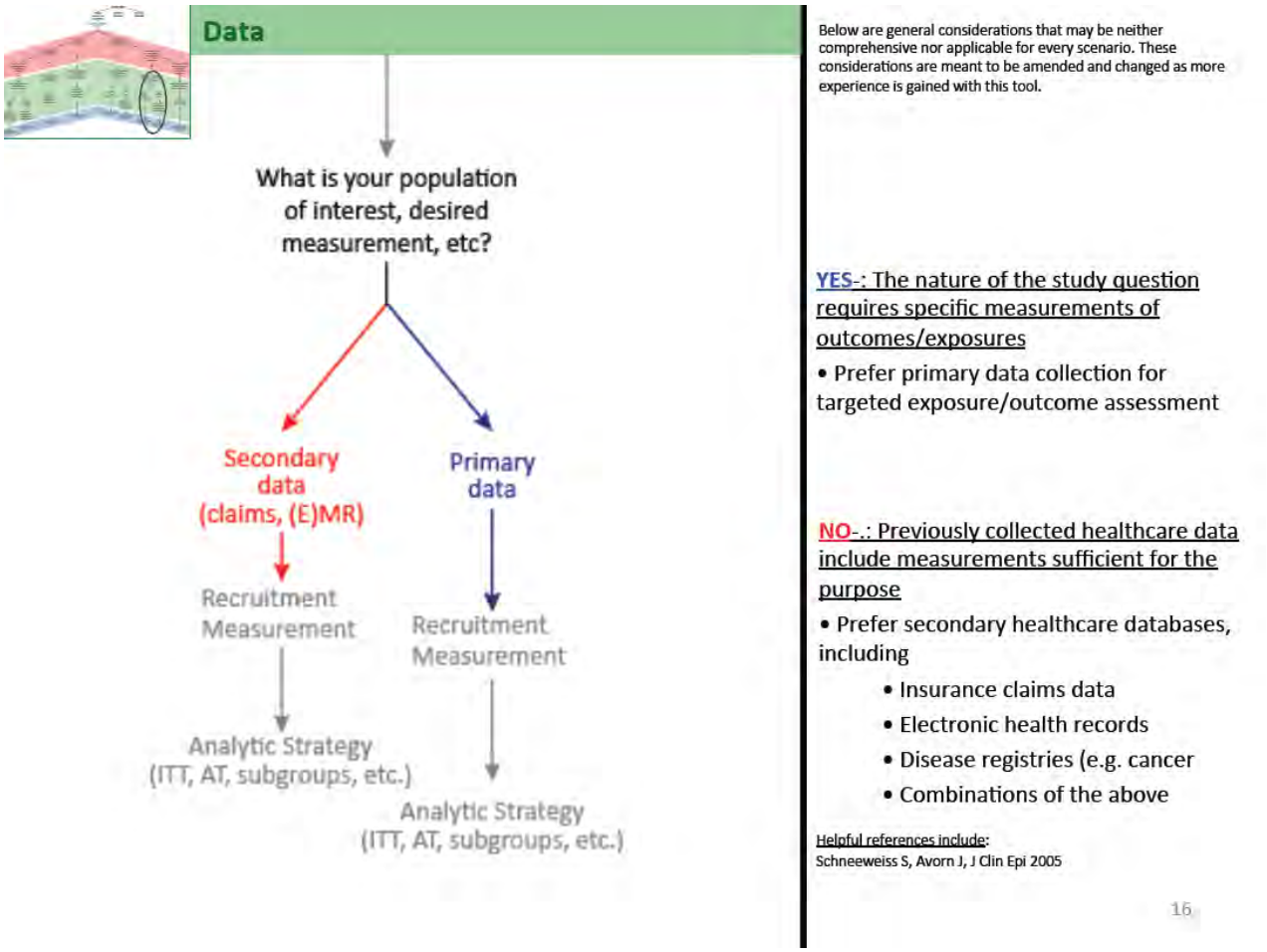
Figure 6.5. Does Exposure Vary Between Patients?



Does exposure vary between patients? In most situations this is answered with a yes. The decision, however, is broader and includes considerations of whether exposure may vary even more on a higher than patient level. Some treatments are subject to strong provider preference to one over another treatment. Such variation can be observed on a regional level, depending on insurance constraints, or over time. Sometimes preference for a treatment changes rapidly after new medical evidence arises. Such variation can be exploited using time trend analyses or instrumental variable

analyses, both of which may provide advantages in confounding control.

Figure 6.6. What is the Population of Interest?



Finally a set of choices needs to be made regarding the most appropriate data sources. Does the nature of the study question require that specific information be newly collected, or can information from previously collected data ("secondary data") suffice to answer the research question? Within the domain of secondary data, several choices are available that need to be considered, including clinical detail, data completeness, access to the data, and confidentiality issues. Often the linkage of multiple data sources is most promising.

Appendix D-4 presents three additional case studies that illustrate the specification of a research question and the thinking that goes into making design, data source, and analytic choices that occur at each of the decision points of a translation tool. The three outlines illustrate the kind of information that the user would enter in to a translation tool.

Conclusions

The Committee hopes that the translation framework proposed here will lead to a series of translation tools appropriate for use by the PCORI Board as it develops funding opportunities in PCOR and to stakeholders for each research category. Because the concept of a translation framework is new, the Committee will be particularly interested in feedback from the Board and public comment on its proposed approach before further developing the framework.

Chapter 7. General and Crosscutting Research Methods

The standards presented in Chapter 7 are those that apply to a broad range of designs. The standards herein are not the only attributes or practices that one would look for when using such approaches, but they represent minimum standards, which if violated could call into question the validity of the research. We have divided these into three categories:

- General methods: Methods that apply to all patient-centered outcomes research, continued from Chapter 3.
- Crosscutting methods: Methods applicable in most or all study designs.
- Data network infrastructure: Methods to develop the sound data networks upon which many types of designs depend.

Methods relevant to specific designs will be covered in Chapter 8.

3.1.3 Identify and Assess Participant Subgroups

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

3.1.4 Select Appropriate Interventions and Comparators

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

7.1.1 Assess Data Source Adequacy

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

7.1.2 A priori, Specify Plans for Data Analysis that Correspond to Major Aims

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

7.1.3 Document Validated Scales and Tests

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

7.1.4 Use Sensitivity Analyses to Determine the Impact of Key Assumptions

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

7.1.5 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 has brought together all reporting guidelines that have been developed using formal approaches, many of which have been adopted by journals, such as CONSORT (for randomized clinical trials), STARD (for diagnostic tests), and STROBE (for observational studies).

Rationale for the Standards

A reasonable starting point for standards for PCOR is the consideration of existing standards for clinical research. The standards selected and proposed in this report are not presented as an exhaustive list of requirements or a how-to manual. Rather, a small subset is presented as core, initial standards to ensure and promote quality PCOR. These standards pertain to planning research, documenting key decisions, and testing the assumptions made in the analyses. They are meant to provide guidance once the researcher has already decided to use the respective methodology. They complement the translation table (Chapter 6), which is intended to shed light on the choices of methodologies and the tradeoff to be made when picking one over another.

High quality, useful research begins with good planning. These planning steps are necessary to ensure that the research will be relevant to actual decisions, that samples and recruitment are adequate (scientific rigor), and that how the research will accomplish its objectives is clear.

Three standards are proposed that highlight PCOR-relevant aspects of planning a high-quality study. In order to ensure the PCOR is relevant to decision-making, investigators should define the clinical decision relevant to a test or treatment that the proposed study is intended to inform. Knowing who will be in the study population is important in knowing to whom PCOR results apply. Another important element in planning a study is selecting appropriate interventions for comparison so that the results of the study reflect the effect that could be expected from actual choices from among treatments used in practice. These treatments should correspond to those defined by the relevant decision specified earlier. Once the clinical decision, the patient sample, and the intervention groups are selected, the relevant outcomes must be defined. When the data include tests or scales, their characteristics as well as evaluations of their performance (psychometric properties) should be established and reported. Next, plans for data analyses should be developed before data are collected that correspond to the major aims of the research. It is important that these methods not just be accurate and appropriate, but also efficient—that they make optimal use of quantitative data so the resulting uncertainty is neither over- nor underestimated.

Once a study is planned, key elements should be assessed and these assessments should be reported. This is subject of three standards. The data to be used for PCOR need to contain all the variables needed for the proposed analyses. This is particularly important in observational studies using pre-

existing data so that confounding can be adequately addressed and outcomes properly measured. Users of the research need to be able to evaluate whether the study produces accurate results and whether the results apply to their situation. This requires that researchers document and report in the protocol key elements of the study design and conduct (e.g., selection of participants, data collection activities, settings, analytic techniques, means of assuring data quality, comparability of study groups, etc.). These standards are essential for transparency and scientific rigor as they allow users of PCOR and other stakeholders to evaluate both the quality of studies and their applicability.

All research requires assumptions about their data analyses for inferences to be valid. Different assumptions have the potential to produce different results. For this reason, assumptions need to be tested to the extent possible, not simply stated. Certain kinds of assumptions, particularly those that are central to an inference and cannot be directly tested using the study data, should be subjected to sensitivity analyses. Sensitivity analyses involve repeating the analyses under different structural assumptions and then comparing the results to see if the conclusions materially change.

Crosscutting Methods

Standards for Causal Inference Methods (See *Standards for Causal Inference* http://www.pcori.org/what-we-do/methodology/)

One of the prime objectives of research is the ability to accurately measure the effect of the actual causes of a health outcome. When that "cause" is a medical intervention, it can be difficult to separate its effects from other factors that might differ among those who had the intervention and those who did not. The randomized trial arose as a methodological answer to this problem: by randomizing participants to one intervention or another, the distribution of independent risk factors of the health outcome, known as potential "confounders" of the causal relationship, was guaranteed (on average) to be equal. This meant that, on average, the estimate of effect of the intervention would be correct.

However, even when trials are executed, many things can happen after randomization—such as informative patient dropout, crossover to other treatments, protocol violations, and missing data—that weaken the power of the original randomization. For some settings and questions, a randomized trial is impossible, undesirable, or simply not done. In such circumstances, we must rely on observational methods—designs in which we observe the outcomes of the interventions that

were decided upon as part of the care process and not by random assignment. However, the complexity and variability of patients and their circumstances, as well as the care they receive, often make it difficult to conclude whether a specific treatment is responsible for the results seen in a research study. Even as diverse study populations and study circumstances that more closely resemble typical clinical care make PCOR more relevant to patients and healthcare providers, these same characteristics make it more difficult to assess the causal impact of a specific intervention on an outcome.

There has long been a suite of tools—most notably various forms of population restriction and regression methods used in virtually all of the sciences—that aim to "control" the effects of confounding variables mathematically, and thereby to produce a valid estimate of an intervention's effect even in complex situations. These methods are both powerful and useful, but they have drawbacks, most notably that they can control only for the effect of confounders that were actually measured. Another, subtler issue is they were not founded on a clearly articulated definition of "cause."

In response to both of these limitations, a new suite of methods have come to be applied in biomedical research called "causal inference methods." Some, such as instrumental variable methods, were borrowed from other fields, like economics, that have used them for decades. Others, called "propensity scores," were first proposed in the 1980s, and have seen much wider use in biomedicine only over the past decade. None of these methods solve the problems of causal inference posed by observational studies, but they can produce estimates of effect and of uncertainty that are more accurate than those from standard methods, partly because they are founded on a more clearly specified conception of "cause" than previous methods.

The Women's Health Initiative (WHI) randomized trial stunned people when results indicated an increased risk of coronary heart disease among postmenopausal women who took estrogen plus progestin hormone replacement. Even as physicians and women around the world switched from embracing to largely shunning hormone replacement, people wondered why the trial results seemed to contradict the findings of multiple large, sophisticated observational studies that had indicated hormone replacement probably offered some protection against heart disease.

Had there been important differences in the observational studies between the groups of women taking hormones and those who were not, differences that went unnoticed? Was it only the randomization used by the WHI researchers that prevented them from being fooled by hidden variables? That was the favored explanation for the startling discrepancy in results.

But sophisticated statistical analysis suggests an alternate explanation: that rather than being tricked by something they failed to measure in the observational data they collected, the researchers were misled by the methods they used to sift through that data.

Other researchers went back to the data of the major observational studies of hormone replacement, but this time they applied techniques that allowed them to mimic central features of a randomized trial, most notably, that subjects were included and analyzed from the point that they newly initiated hormone therapy. Even though they used the same observational data, this analyses produced different results– results that closely matched the findings of the WHI trial. Their reanalysis, using better methods, pointed to differences in the time that passed since menopause, time patterns of risk after estrogen initiation, and different durations of follow-up as the reason the observational trials had not accurately foreshadowed the WHI results.

Source: (36)

It should be noted that propensity scores, like standard regression methods, cannot solve the problem of unmeasured confounding factors, but they can allow adjustment for more confounders or proxies of unmeasured confounders. (See Appendix B from Rosenbaum PR, Rubin DB. Reducing bias in observational studies using subclassification on the propensity score. Journal of the American Statistical Association, 1984, 79(387):516-524.) Instrumental variable methods, on the other hand, do purport to get around the unmeasured confounder problem by identifying and exploiting naturally occurring quasi-random treatment choices in the healthcare system, but they do so at the expense of relying on assumptions that are untestable using the data at hand, and these assumptions require extraordinarily close scrutiny.

Even though the methods and literature about causal inference are in their early stages, they are being used with increasing frequency in PCOR. One reason for this is that many are turning to electronic healthcare databases to quickly enable such studies in large populations in standard care settings as opposed to research settings. But few of those studies are randomized, and thus observational methods are being increasingly relied upon to extract estimates that can support causal interpretations from data that may not have been produced or gathered with such an intended use.

Causal Inference Standards

7.2.1 Define Analysis Population Using Information Available at Study Entry

Decisions about whether patients are included in an analysis should be based on information available at each patient's time of study entry and not based on future information, such as future changes in exposure.

7.2.2 Describe Population that Gave Rise to the Effect Estimate(s)

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

7.2.3 Precisely Define the Timing of the Outcome Assessment Relative to the Initiation and Duration of Intervention

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

7.2.4 Measure Confounders before Start of Exposure

In general, variables for use in confounding adjustment (either in the design or analysis) should be ascertained and measured prior to the first exposure to the therapy (or therapies) under study.

7.2.5 Assess Propensity Score Balance

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

7.2.6 Assess Instrumental Variable Assumptions

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

Rationale for the Standards

In general, because they are not based on randomization, observational studies must pay substantial attention to and protect from sources of bias. For example, most researchers agree that all patients included in a randomized controlled trial should be analyzed with their assigned group, regardless of

any changes that occur during the study. This is referred to as intention-to-treat analysis. Observational studies can take an equivalent approach by including patients based on factors known at the defined start point and not based on later changes (Standard 7.2.1). An obvious starting point is to specify who is included in the analysis, why they are included, whether any variables measured after baseline may introduce bias, and how the different groups compare on key characteristics. In order to increase the accuracy of results, researchers may include only selected patients in some analyses. For example, patients might be separated by age or the severity of their illness. In some cases, statistical methods such as propensity scores can combine several characteristics into one variable that is used for matching.

If the selection process means that the study population no longer adequately resembles the sort of patients who face decisions about the treatment or service being studied, or if the selected group of study participants is too small, the research may not produce convincing or relevant results. For this reason it is important that researchers describe which patients were included in any analysis and why they were selected. When a sophisticated analytical approach such as propensity scores is used, additional effort is required to demonstrate that the resulting study groups are balanced on key characteristics.

Measuring and adjusting for pre-treatment 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, researchers may introduce bias, if these variables are affected by the study treatment. Such bias may make it harder to be certain the treatment is causing the result.

Standards for Studying Heterogeneity of Treatment Effect (See *Standards in Addressing Heterogeneity of Treatment Effectiveness* http://www.pcori.org/what-we-do/methodology/)

Heterogeneity of treatment effect (HTE) is a technical term for the phenomenon that people can respond differently to the same treatment. In some people the treatment will produce the intended benefit, in other people there may be some benefit but less than intended, while in others the treatment may have no effect or even cause harm. In traditional clinical research this variability can be masked by the study design and analysis, by not measuring the variables that predict different responses, or by not analyzing them. In both clinical trials and observational studies, results can be

averaged across all the patients in a study, obscuring how responses to a treatment might vary within the study population. As a result, it can be hard to determine from research results what the results of a treatment will be for a specific type of patient.

Explicitly addressing HTE in research helps answer the question "What is likely to happen to patients like me?" and makes research results more useful for patients and clinicians who need to make decisions about the best course of treatment for specific patients.

HTE analyses can contribute to the goals of PCOR by either 1) estimating the effects of a treatment in subgroups of patients or 2) predicting whether a specific person will benefit from treatment. The most common approach is to use subgroup analyses to estimate the effects of treatments in patient subgroups. Prediction of individual effects is less common, though decision analysis and microsimulation models have been used to predict patient benefit. The proposed standards apply to estimating subgroup treatment effects; future PCORI work may consider standards for decision analysis and microsimulation modeling.

Designing studies and conducting analyses that provide accurate answers about heterogeneity of treatment effects among patient subgroups is challenging. Investigators generally use two different approaches to estimate heterogeneity of treatment effect: they 1) estimate the effect of treatment separately for patient groups or 2) use statistical modeling to determine the interaction between the treatment and patient characteristic. Estimating heterogeneity of treatment effect by stratifying by subgroup (e.g., men vs. women) is susceptible to well-established problems that can result in increased likelihood of falsely detecting HTE (Type I error) or failing to detect true HTE (Type II error), when compared with examining whether the treatment works, on average, in the overall sample. While stratified analysis is valid for estimating stratified treatment effects, it is incorrect for inferring HTE. To correctly test for HTE, investigators need to test whether the difference between the two stratified treatment effects is zero using an interaction test, assuming the variability is similar between the treatment groups. In a randomized trial, patients are randomized to the intervention; subgroups are not randomized. Subgroups may have different baseline characteristics, confounding the interpretation of results. In addition, testing for HTE with multiple subgroups results in multiple comparisons, increasing the likelihood of finding HTE when none exists. Use of subgroups also divides the sample, reducing statistical power.(37, 38) Interaction tests require fewer statistical tests and have more statistical power than do subgroup analyses. Brookes(39) showed that when

stratifying by subgroup, there is up to 66 percent probability of inferring the presence of HTE when the overall treatment effect is significant, and that there is up to 25 percent probability of inferring HTE when the overall treatment effect is non-significant, even when there is no actual HTE.

Subgroup analyses may be undertaken for different reasons, including hypothesis testing, HTE estimation, and descriptive purposes. Table 1 describes a framework endorsed by PCORI that describes the characteristics of analyses for these different purposes. Minimum standards are proposed according to this framework for inferential goals of HTE analysis in patient-centered outcomes research.

The consideration of subgroup analyses for descriptive purposes is less well established due to concerns about misinterpretation of descriptive subgroup analyses (i.e., publishing comparisons by sex that do not show a significant difference). Some have asserted that if a study is not adequately powered to detect subgroup effects, subgroup analyses are meaningless. However, descriptive reporting of subgroup effects is important in synthesizing subgroup results in a meta-analysis or in using Bayesian inference methods when no individual study is powered to detect subgroup differences. The U.S. Food and Drug Administration and the *Journal of the National Cancer Institute* both recommend reporting of treatment by sex and other subgroups.(40, 41) In addition to sex, other subgroup variables that may be important to report include demographic characteristics (e.g., age), behavior (e.g., smoking), pathophysiology (e.g., measures of disease severity), genetic markers, and comorbid conditions (e.g., diabetes status in cardiovascular disease trials).

Table 7.1. The Essential Characteristics of the Three Different Types of HTE Analyses

Properties	Confirmatory HTE Analysis	Descriptive HTE Analysis	Exploratory HTE Analysis
Inferential goal	To test hypotheses related to subgroup effects	To report treatment effects for future synthesis	To generate hypotheses for further study
Number of subgroups analyzed	A small number, typically one or two	Moderate to large	May be large
Scientific rationale and prior evidence for hypotheses	Strong	Immaterial	Weak or none
Pre-specification of data analytic strategy	Fully pre-specified	Fully pre-specified	Not pre-specified
Control of family-wise type I error probability or qualitative interaction	Should be done	Not needed	Difficult, because it is not obvious how many related tests were performed
Characterization of sampling properties of the statistical estimator (e.g., standard errors, type I error rate, coverage probability)	Easy to achieve	Possible	Difficult
Power for testing hypothesis	Ideally, study designed to have sufficient power	Likely to be inadequately powered, but this is immaterial	Typically, inadequate power to examine several hypotheses

Abbreviations: HTE=heterogeneity of treatment effect.
(Adapted from *Standards in Addressing Heterogeneity of Treatment Effectiveness* http://www.pcori.org/what-we-do/methodology/.)

Standards for Heterogeneity of Treatment Effect (HTE)

7.3.1 State the Goals of HTE Analyses

State the inferential goal of each HTE analysis; identify each analysis as confirmatory, descriptive, or exploratory. See Table 7.1 comparing the different types of HTE analyses.

7.3.2 For Confirmatory and Descriptive HTE Analyses, Pre-specify Subgroups and Outcomes; for Confirmatory HTE Analyses, Pre-specify Hypotheses for Each Subgroup Effect

The study protocol should unambiguously pre-specify planned confirmatory HTE analyses. Pre-specification of confirmatory HTE analyses should include a public record with a clear statement of the hypotheses the study will evaluate, including the definitions of subgroup variables and outcomes, and the direction of the expected treatment effects. Prior evidence should be available for review and the study protocol should present this evidence clearly. The hypotheses for descriptive HTE need not be pre-specified; rather, specify the subgroups to be studied, as one goal is to facilitate future meta-analyses.

7.3.3 For Confirmatory HTE Analyses, Report a priori Statistical Power

Studies should calculate and report the power to detect treatment effects in each subgroup and to detect the interaction between the treatment and the subgrouping variable (i.e., the power to test whether the effects are statistically different between particular subgroups).

7.3.4 For Any HTE Analysis, Perform an Interaction Test and Report Sufficient Information on Treatment Effect Estimates

To detect differences in treatment effect between subgroups, use an interaction test (i.e., test whether the interaction between the treatment indicator and the subgroup variable is statistically significant).

Within each subgroup level, studies should present treatment effect estimates, standard errors, and 95 percent confidence intervals. Studies should also report the P-value for the interaction test for each subgrouping variable. For descriptive analyses, studies should also consider presenting a forest plot as a visual summary of the results, although such forest plots should not be used to infer HTE.

7.3.5 For Exploratory HTE Analyses, Discuss Findings in the Context of Study Design and Prior Evidence

Exploratory HTE analyses should be presented in the context of whether they are consistent with prior evidence and how well the study design addresses the HTE question. These considerations are more important than P-values for inferences.

7.3.6 For Any HTE Analysis, Report All Pre-specified Analyses and, at Minimum, the Number of Post-hoc Analyses, Including Number of Subgroups and Outcomes Analyzed

Studies must report the results of all the HTE analyses that were pre-specified in the study protocol or grant application, regardless of their statistical significance. Reports of exploratory HTE analyses that did not pre-specify subgroups should clearly report the number of subgroups and outcomes analyzed.

"There is a crash making a mess of traffic on westside freeway. All southbound lanes are blocked so a Healthflight helicopter can pick up a seriously injured victim."

It's a common decision faced by first responders: transport a patient by ambulance or call in a helicopter. The decision is complex, involving distance, injury severity, and more. Helicopters not only cost more, but they often cause bigger traffic jams and other disruptions.

But then there's the ultimate question: will calling in a helicopter save a life?

No one has seriously proposed randomly assigning crash victims to air or ground transport. Like many other vital questions about healthcare choices, answers will have to come from observations of everyday practices.

Several studies have given the nod to helicopters. But hang on—there might be cracks in the foundation of evidence here. Trauma scenes are chaotic. Decisions are made quickly. Blank spots in records of vital signs, severity of injuries, distance, and transport time make it difficult to compare patients who flew with those who went by ground. What's more, since patients were not randomly assigned to air or ground transport, researchers have to use other methods of comparing groups of patients.

A recent study concluded once again that helicopters may offer some advantage in some cases. For every 65 flights, one patient with major trauma survived who might not have lived if she had gone by ground ambulance. The researchers caution they don't know if there are differences in helicopter and ground ambulance crews or other factors besides the helicopter itself that could explain their results.

But they note that the apparent advantage they calculated was slimmer than that reported by earlier studies. At least some of the gap seems to be connected to how the different studies dealt with the inevitable blank spots in medical and other records they relied on. But they also noted that much of the apparent benefit of helicopter transport might really be due to differences in the patients. Using what's called a propensity score, these researchers calculated the likelihood a patient would be sent by air or ground. That sophisticated statistical method led to a more conservative estimate of any advantage to using more costly and more disruptive helicopters, thus giving policy-makers a clearer view of their value compared with other potentially lifesaving investments.

Source: (42)

Rationale for the Standards

Given the importance to patients, providers, and stakeholders of subgroup analysis for understanding heterogeneity of treatment effect, it is important that these analyses are valid. The first step is to state the goal of the analysis; this will direct the appropriate design and analysis plan for the study and also allow stakeholders to correctly interpret results.

Specification of subgroups and reporting the number of subgroups tested ensures that methods are transparent and errors from multiple statistical comparisons (e.g., Type I or S errors or finding a difference when no difference exists) are detected. For confirmatory HTE analysis, it is important to calculate and report a priori statistical power. This standard promotes scientific rigor as it ensures that investigators examine the prior evidence to collect information on plausible effect sizes and their variation across subgroups and allows readers to assess the reliability of study findings, especially when no HTE was detected.

Standards for Preventing and Handling Missing Data (See *Minimal Standards in the Prevention and Handling of Missing Data* http://www.pcori.org/what-we-do/methodology/)

Missing data are unrecorded or unavailable data values that would be meaningful for the analysis. Missing data, if allied with the use of improper statistical methods for handling missing data, can bias the results or overstate their precision.

Missing data are virtually inevitable in studies of humans. It occurs when data fail to be recorded because of patient actions, such as study dropout, failure to return for follow-up, or unwillingness to provide certain data; by design, when certain measurements are not planned to be recorded; or simply from errors in data measurement or recording. Data sets derived from records not intended for research, such as those generated from routine clinical care, are particularly prone to having missing data, as are those that involve patient populations that are harder to retain in studies. When patient-centered outcomes research takes place in realistic care settings and includes a broad variety of patients, researchers are more likely to encounter challenges with missing data. Patients with multiple disease conditions and those seen in community care settings may be more likely to be lost to follow-up. Often, to prevent missing data, researchers conduct studies in specialized clinical settings and exclude patients who, because of other clinical problems, might be less likely to complete the study. As a result, the research may fail to represent actual results that would occur in more varied clinical settings and more diverse patient groups.

In order to produce results that are most likely to be both accurate and relevant to patients, researchers must follow the best available practices for both minimizing and managing the impact of missing data.

Standards for Preventing and Handling Missing Data

7.4.1 Describe in Protocol Methods to Prevent and Monitor Missing Data

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

7.4.2 Describe Statistical Methods to Handle Missing Data in Protocol

Statistical methods for handling missing data should be pre-specified in study protocols, and their associated assumptions stated in a way that can be understood by stakeholders. The reasons for missing data should be considered in the analysis. This standard applies to all study designs for any type of research question.

7.4.3 Use Validated Methods to Deal with Missing Data that Properly Account for Statistical Uncertainty Due to Missingness, Such as Multiple Imputation; All Forms of Single Imputation Are Discouraged

Statistical inference of intervention effects or measures of association should account for statistical uncertainty attributable to missing data. This means that methods used for imputing missing data should have valid type I error rates and that confidence intervals should have the nominal coverage properties. This standard applies to all study designs for any type of research question. Multiple imputation methods satisfy this condition, along with various likelihood-based and other validated methods. Single imputation methods like last observation carried forward and baseline observation carried forward are discouraged as the primary approach for handling missing data in the analysis.

7.4.4 Record and Report All Reasons for Dropout and Missing Data, and Account for All Patients in Reports

Whenever a participant discontinues some or all types of participation in a research study, the investigator should document the following: 1) the reason for discontinuation; 2) who decided that the participant would discontinue; and 3) whether the discontinuation involves some or all types of participation. Investigators should continue to collect information on key outcomes for participants who discontinue their protocol-specified intervention. This standard applies to all prospective study designs that aim to assess intervention effectiveness.

All participants who enter the study should be accounted for in the report, whether or not they are included in the analysis. Describe and justify any planned reasons for excluding participants from analysis. This standard applies to all study designs for any type of research question.

Standards for Preventing and Handling Missing Data (continued)

7.4.5 Examine Sensitivity of Inferences to Missing Data Methods and Assumptions, and Incorporate into Interpretation

Examining sensitivity to the assumptions about the missing data mechanism (i.e., sensitivity analysis) should be a mandatory component of the study protocol, analysis, and reporting. This standard applies to all study designs for any type of research question.

Tables should be used to describe missing data studies. Potential bias resulting from imperfect definitions should be discussed with an estimate of the change in the direction and magnitude of the effect due to bias. These quantitative results should be incorporated into the interpretation of the study and reflected in the discussion section and possibly the abstract. If there are big effects, help the user further understand the reason for the missing data and the effect of the missingness on the findings.

Rationale for the Standards

Preventing and planning for missing data, and describing the methods used for it in a study protocol, are minimal requirement for good research. Tracking all patients in a study and recording reasons for dropout and loss to follow-up is currently considered good practice and is required by many of the organizations that fund research and the journals that report results. The extent and pattern of missing data must be clearly reported.

The use of suboptimal or even invalid methods for analyzing datasets with missing data is widespread in the literature. The reasons for this are unclear, although it is likely that the ease of implementation of inappropriate methods and perhaps a lack of awareness among researchers of their limitations has contributed. The tendency for researchers within a field to judge the appropriateness of the methods they use by looking to the methods used previously by other scientists within that area and the willingness of journals to publish papers using those methods results in a resistance to change within research specialties. Many researchers and groups have attempted to improve this situation(43) but it is clear that the health research community requires further and firmer guidance. The science and software are sufficiently mature that there is no longer any reason to use improper methods to handle missing data.

Determining how the analysis will address missing data before seeing the data is not always common practice, but it reduces the chance of selecting an inappropriate approach. In the past 30 years, many

Imagine you are collecting data on an extremely rare drug side effect, a side effect so rare that in a database of 200,000 patients (half taking the drug and half not) there is not a single instance of the side effect in either group. But then, just by chance, one patient experiences the side effect in a small trial with just 100 patients. Would you ignore the previous experience and conclude the risk is 1 percent based on the new, small study? Of course not. This imagined scenario is extreme, but it is common practice when performing meta-analyses to exclude datasets where no events were recorded.

A prominent meta-analysis of cardiovascular risk associated with the use of the diabetes drug rosiglitazone (Avandia) concluded the drug raised heart attack risk by 43 percent and cardiovascular death risk by 64 percent. The meta-analysis used data from 42 clinical trials, but excluded four trials from the heart attack risk analysis and 19 trials from the cardiovascular death risk analysis because no relevant events were recorded.

A reanalysis using more sophisticated methods to incorporate all of the available data could not establish an association between rosiglitazone and cardiovascular risk. The authors of the reanalysis wrote: "We think that excluding trials with zero events in the index meta-analysis probably exaggerated risk estimates and that including these trials by applying continuity adjustments in this instance temper [sic] the exaggerated estimates."

Source: (44)

new methods for handling missing data have been developed, and some may require more expertise to use. Methods that use multiple values for the missing value are more likely to produce accurate results and should be used in most situations rather than using either a single value (e.g., the baseline or the last observation carried forward) or only including cases with complete data.

All missing data methods rely on assumptions that cannot be assessed directly in the dataset. The three main assumptions are that the chance something is missing has nothing to do with a patient's characteristics (known as "missing completely at random"); that it depends on patient characteristics predictive of the outcome, but ones that were measured ("missing at random"); and finally, that missingness is dependent on patient characteristics predictive of the outcome that were either not measured or observed ("missing not at random," or "non-ignorable" missingness). The sensitivity of inferences to those assumptions should be routinely assessed if the degree of missing data is likely to materially affect either the bias or precision of an effect estimate.

Data Networks as Research-facilitating Infrastructures (See *Standards in the Use of Collaborative or Distributed Data Networks* http://www.pcori.org/what-we-do/methodology/)

Collaborative data networks are agreements that facilitate the use of data across organizations and/or locations for healthcare research. Data networks are used to aggregate information from a range of data sources (claims, medical records, lab/pathology reports, etc.) or from various locations (health plans, hospitals, clinics, care facilities, etc.) for use by multiple research studies on various topics.

Key elements of a data network as a research-facilitating infrastructure include an architecture (structure) that allows networking, privacy policies that protect patient information, governance guidelines that specify roles and responsibilities, and rules for how data elements are defined, described, and organized. These key elements create the infrastructure for data networks; they do not determine the research questions or the research design. Networks may be used to establish disease-specific registries, create broad-ranging surveillance systems, or facilitate the conduct of randomized trials. They may cover a wide range of research topics, including but not limited to studying the effectiveness of diagnostic tests, monitoring the risk for adverse effects of new drugs or devices, and testing new cancer treatments.

Data networks are important research infrastructures for the development and advancement of PCOR. Analyzing already-collected data across organizations or locations is more efficient than replicating studies in multiple locations or populations. Studies based on networked data are also likely to include more types of patients and variations in treatment patterns than would be available in any one site. This means the results are more likely to be generalizable—useful to more patients and clinicians. Almost by definition, data networks include larger numbers of patients than can normally be enrolled in most trials and cohort studies. While numbers alone do not improve a study, they do increase the precision in effect estimates to detect smaller differences in outcomes or to detect differences more quickly after a new medical product is marketed. Equally important for PCOR, large numbers allows analyses of treatment effectiveness in larger numbers of patient subgroups. With large numbers of records, it is easier to determine whether the comparative effectiveness observed across all study participants is consistent or varies across meaningful subgroups (e.g., between men and women or among people with different comorbidities).

A data network is only as good as the quality of its data. The challenges in establishing and maintaining data networks include harmonizing both technical aspects of databases and the expectations and responsibilities of the participating organizations. Definitions and other characteristics of data elements need to be clear, agreed-upon, and verified. Creating and maintaining standardized terminology and descriptions of the data requires planning and resources. Collective agreement and clarity is also needed about how patient privacy will be protected, who has access to the data, and who owns both the data and the results of the research.

Data Networks as Research-facilitating Infrastructures

7.5.1 Data Integration Strategy

In order for equivalent data elements from different sources to be harmonized (treated as equivalent), processes should be created and documented that either a) transform and standardize data elements prior to analysis or b) make transformation logic available that can be executed when data are extracted. The selected approach should be based on an understanding of the research domain of interest.

7.5.2 Risk Assessment Strategy

Data custodians should assess the uniqueness of records (i.e., no other records have the same values) of patient records to measure re-identification risk of data, and apply algorithms to ensure that the desired level of confidentiality is achieved to meet the need of the particular PCOR application.

7.5.3 Identity Management and Authentication of Individual Researchers

Develop a reliable process for verifying credentials of researchers who are granted access to a distributed research network and authenticating them.

7.5.4 Intellectual Property Policies

A research network should develop policies for the handling and dissemination of intellectual property (IP); networks should also have an ongoing process for reviewing and refreshing those policies. IP can include data, research databases, papers, reports, patents, and/or products resulting from research using the network. Guidelines should balance (1) minimizing impediments to innovation in research processes and (2) making the fruits of research widely accessible, particularly to the people who need them the most.

7.5.5 Standardized Terminology Encoding of Data Content

The data contents should be represented with standardized terminology systems to ensure that their meaning is unambiguously and consistently understood by the party using the data.

7.5.6 Metadata Annotation of Data Content

Semantic and administrative aspects of data contents should be annotated with a set of metadata items. Metadata annotation helps to correctly identify the intended meaning of a data element and facilitates an automated compatibility check among data elements. A data element is the entity and its property described with a given data. A data element is defined with a set of metadata.

7.5.7 Common Data Model

Individual data items should be assembled into a contextual environment that shows close or distant association among data. A common data model (CDM) specifies necessary data items that need to be collected and shared across participating institutes, clearly represents these associations and relationships among data elements, and promotes correct interpretation of the data content.

Rationale for the Standards

Several organizations in the U.S., Canada, and Europe have developed guidelines, best practices, and initiatives for defining key elements 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 healthcare, like HIPAA in the U.S., or general with applications in healthcare, such as the Organization for Economic Co-operation and Development personal privacy guidelines.(45) Compiling all of these would result in a lengthy technical manual that could be prescriptive and potentially hamper innovation in what is a rapidly evolving and growing field. Selected minimum standards are proposed to ensure the essential elements of data quality, privacy, and collaboration.

Constructing data networks so that the quality of data is defined improves the scientific rigor of the studies that use the data network. Clearly specifying how this is done also promotes transparency by allowing others to evaluate the methods and replicate them. For a data network to function and provide useful data, processes should be created and documented that transform data elements so they are equivalent, terminology should be standardized, and information about the data elements, called metadata, needs to be provided. These data elements also need to be assembled into a model that shows the relationships among the data elements and helps all users to interpret the data correctly.

Data networks link and share individual-level data in ways that could compromise patient privacy. Generally, PCOR proposals and protocols should describe data use agreements, informed consent, data security, and approaches to protecting security. The proposals should also describe how these address the risk of re-identification of patients and the actual use of data compared with the originally designed and consented use of the data. In order for patients and clinicians to realize the benefits of research using data networks without the risk of harms that could result if privacy is not maintained, standards are required to limit and control who has access to the data. Additionally, data networks need to be proactive and evaluate if any use or characteristic of the structure of the network is likely to compromise confidentiality and address these issues.

Finally, the utility of a data network often increases with the longevity of the network. Longevity requires that the participating organizations maintain relationships and continue to collaborate over time. 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 any products resulting from the network.

Recommended Actions

In order to support use of the general and crosscutting standards discussed in this chapter, PCORI should:

- Sponsor randomized trials alongside registry studies to compare the validity of different methods for reducing confounding and bias.
- Develop and disseminate software needed for sensitivity analyses and approaches to evaluating the assumptions underlying complex analyses such as instrumental variable analyses.
- Develop and distribute software to reduce barriers that inhibit the use of more rigorous methods for handling missing data.
- Provide training in methods for systematic reviews, modeling, and addressing missing data.
- Promote accumulation of evidence to supplement common practices as guidance for future development of data network structures.
- Promote approaches to privacy protection in data networks that also consider how to enhance data utility.
- Require that missing and loss-to-follow-up data be reported in all research results.

Research Recommendations

- Fund research on innovative ways to identify and recruit new users of interventions for research studies.
- Fund research on ways to identify and include reasonable treatment alternative comparators.
- Develop and disseminate templates for describing who is in each analysis and the potential sources of selection bias.
- Develop and disseminate methods for adequate analysis of data in cases where the treatment/exposure varies over time and it is not possible to adhere to these standards.
- Incorporate evolving new technology, such as the use of cloud technology, into ongoing work in the design of networks.
- Fund research on the best way to harmonize data elements across sources.
- Develop methods guidance for analyses for HTE in observational studies.
- Develop methods guidance on the use of Bayesian methods in HTE analyses and appropriate outcome scale for HTE analysis (e.g., risk difference, risk ratio, log of odds-ratio).
- Support the development of both analytic approaches and guidance for predictive approaches to HTE as well as for SGA with a focus on their use for PCOR.
- Develop methods guidance for HTE analyses in comparative effectiveness trials; the literature on HTE almost exclusively discusses use in placebo controlled trials.

Chapter 8. Design-specific Methods

Standards for Adaptive and Bayesian Trial Designs (See *Standards for the Design, Conduct, and Evaluation of Adaptive Randomized Clinical Trials* http://www.pcori.org/what-we-do/methodology/)

Randomized trials can provide the strongest evidence about the comparative effectiveness of different treatments, but they are often perceived to take too long or be too rigid in a rapidly changing world. Adaptive trials allow changes to be made to a study while it is ongoing. Examples of "adaptations" include changing what proportion of patients are randomized to which group, altering the sample size, changing the eligibility criteria, dropping or adding comparison arms, changing endpoints, and stopping early. Rather than waiting until the end of the study period to implement changes, the changes are planned for as part of the trial design and executed based on the analyses conducted during the trial.

Recognizing the need for innovation in clinical trial design, representatives from the NIH's Clinical and Translational Science Award programs have identified adaptive clinical trial design as a high-priority methodological issue "to increase the efficiency of comparative effectiveness trials."(1) Adaptive designs are particularly appealing for PCOR because they have the potential to maintain many of the advantages of randomized clinical trials while minimizing some of the disadvantages. Adaptive trials can sometimes provide faster results, and can also increase the relevance of trial results by adjusting both the composition of patient groups and the treatments based on interim results and clinical questions. The flexibility and efficiency that are gained in adaptive trials have to be balanced with the risk that such trials typically require a longer design period and involve more logistical complexity. Also, there are relatively few statisticians with expertise or experience in designing or carrying out such trials.

Adaptive designs for trials are not new, but they have gained in popularity in recent years due in part to efforts to streamline drug and device development. To date, the use of adaptive trials for PCOR has been limited, but given the potential, researchers are looking for guidance on how to design and conduct PCOR adaptive trials. A literature search found only two adaptive trials that could be considered PCOR, one comparing insulin regimens and another of chemotherapy in older women with early-stage breast cancer.(46, 47) That said, many trials have some adaptive features; stopping

guidelines and sample size re-estimation are part of standard practice. Many adaptive features can be implemented individually using frequentist approaches, but complex designs combining several dimensions of adaptation typically require a Bayesian approach.

Standards for Adaptive and Bayesian Trial Designs

8.1.1 Specify Planned Adaptations and Primary Analysis

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

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

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

8.1.2 Evaluate Statistical Properties of Adaptive Design

While not necessary for simple designs, the statistical properties of complex adaptive clinical trial designs should be thoroughly investigated over the relevant range of important parameters or clinical scenarios (e.g., treatment effects, accrual rates, delays in the availability of outcome data, dropout rates, missing data, drift in participant characteristics over time, subgroup-treatment interactions, and/or violations of distributional assumptions). Statistical properties to be evaluated should include type I error, power, and sample size distributions, as well as the precision and bias in the estimation of treatment effects. Additional performance metrics may also be evaluated (e.g., the frequency with which specific adaptations occur, the likelihood of substantial covariate imbalance, the likely adequacy of final data for subgroup and safety analyses).

The programming code used to create the simulations should be retained with version control. The programming code and software used should be made available to stakeholders who have a need to know, including reviewing agencies.

8.1.3 Specify Structure and Analysis Plan for Bayesian Adaptive Randomized Clinical Trial Designs

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

8.1.4 Ensure Clinical Trial Infrastructure Is Adequate to Support Planned Adaptation(s)

The clinical trial infrastructure, including centralized randomization, data collection related to the assessment and recording of key outcomes, data transmission procedures, and processes for implementing the adaptation (e.g., centralized, web-based randomization), must be able to support the planned trial. In simple adaptive trials, qualitative verification of the capabilities of the proposed trial infrastructure may be adequate. Trials with more complicated requirements such as frequent interim analyses require thorough testing prior to trial initiation. Such testing should involve the trial's data collection and data management procedures, the implementation of the adaptive algorithm, and methods for implementing the resulting adaptation(s). The impact on the trial's operating characteristics of delays in collecting and analyzing available outcome data should be assessed.

8.1.5 Use the CONSORT statement, with Modifications, to Report Adaptive Randomized Clinical Trials

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

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

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

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

Rationale for the Standards

While current practice does not provide extensive guidance for adaptive trials in PCOR, the experience in therapeutics and device trials, combined with theoretical considerations, allows us to identify basic rules governing design and conduct.

Adaptive trials should adhere to the principles of good design and analysis that apply to all rigorous research. The complexity of adaptive trials can make this more difficult, requiring additional attention to specific steps in the research process. For example, these studies typically require simulations in the design phase in order to define the error rates, and descriptions of the design both in protocols and published papers must include more elements than a non-adaptive trial. Good adaptive trial design requires pre-planning and specification of procedures at the outset. Given the potential complexity introduced by adaptations, the timing of interim analyses and the changes that could be made based on those data should be determined before the trial starts. Similarly, standardized reporting of trials has become part of best practice and, to the extent that existing reporting guidelines (i.e., CONSORT) can be used, they should be followed and any modifications described.

Other components of adaptive trials require special focus. Adaptation requires an infrastructure to obtain and analyze the data needed for design changes as the trial proceeds. This capacity is not the norm in conventional trials and demands special attention.

Adaptive trials that use Bayesian approaches require even more detailed specification of the analysis plan than is current practice or would be required in traditional trials, both because software is not standardized and because Bayesian methods have analytic features not present in standard trials.

Standards for Data Registries (See *Standards in the Conduct of Registry Studies* http://www.pcori.org/what-we-do/methodology/)

A registry is an organized system that collects data for scientific, clinical, or policy purposes, and as such, forms the basis of an observational study. Evaluation of the quality of registries involves assessing the design, data elements and sources, and governanceRegistries are structured systems for collecting and organizing uniform data about the progress and outcomes associated with the course of disease or associated with the defining characteristic of the registry (e.g., familial cancer risk or

device implantation). They are typically single cohorts that are assembled in anticipation of future research related to their focus. When questions arise that can be answered with such data, answers can often be obtained quickly because of the comprehensiveness of risk and outcome data already in the registry. However, because registries may not gather all the information needed for certain questions that arise after their inception, because they can be affected by a variety of time trends, and because they typically do not include control populations, special attention needs to be paid to issues of data quality and biases in studies that utilize them.

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

Registries are particularly relevant to and important for PCOR. When properly designed, they are able to provide data on groups of patients sometimes not included in clinical trials, and they can be 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 actually delivered in "real world" situations increases the likelihood that the findings will be broadly applicable to many people and situations.

Successful joint replacement surgery can return a patient's mobility, relieve pain, and improve quality of life. Replacement joints are medical devices, and their safety and efficacy are tested in carefully controlled clinical studies before they are approved for general use. However, clinical trials cannot test artificial joints in all types of patients likely to use them, nor can studies be continued for decades, which is how long patients are likely to live with an artificial joint. Registries provide a way to collect the data needed to track the outcomes of surgery and the performance of different types of artificial joints over an extended period.

In 2010, analysis of data collected in the National Joint Registry of England and Wales revealed that patients with metal-on-metal hip implants required a second surgery to replace or remove part of the artificial joint (called a revision) more frequently than was expected based on comparisons with other types of artificial joints. This led to further investigation, a worldwide recall of one metal-on-metal implant by the manufacturer, more intensive monitoring by the FDA in the U.S., and research about metal-on-metal joints using other joint registries maintained by provider organizations, U.S. States, and other countries.

The same characteristics of registries that make them reflective of real-world practice sometimes limit their usefulness in informing healthcare decisions. This is where methodological standards can help. Data generated in the "real world" are "messy" (i.e., not as tightly controlled as in a clinical trial or even some prospective cohort studies), and definitions may differ across data sources and change over time, making it difficult or impossible to design unbiased studies and draw proper conclusions. Careful planning prior to establishing a registry or beginning a registry study can help minimize data variability and improve study results. In addition, without careful planning and oversight, there can be problems with the use and confidentiality of the data entered in registries. Also, tracking and matching patients across sources and over time is resource-intensive and can result in the identification of patients unless carefully planned. Since registries typically follow the natural history of patients, they require multiple points of follow-up. Registries are often most useful when they are maintained over long enough periods to obtain important long-term outcomes. The need for recurrent patient contact over time presents the problem of significant quantities of missing data since the data may not be available or the patient cannot be reached. Perhaps most importantly, since patients are not randomized and may end up in the registry for a variety of reasons, registry studies require careful consideration of various biases that might exist. Confounding is particularly a concern, and research based on registries must contain data elements that will allow for statistical controls for confounding, while researchers must develop complementary approaches to analysis. If the data collected are not standardized (e.g., definitions and follow-up), large amounts of missing data exist, or confounding is not controlled, wrong conclusions can be drawn which could do more harm than good.

Standards for Data Registries

3.1.1 Develop a Formal Study Protocol

In addition to the general study protocol standards, registry studies should include a formal study protocol specifying: at least one purpose of the registry (e.g., effectiveness, safety, natural history of disease, quality improvement, or other); data sources and linkage plans, if any; measure(s) of effect; and use of any standardized data dictionaries (nationally or internationally accepted).

3.1.5 Measure Outcomes that People in the Population of Interest Notice and Care About

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

8.2.1 Describe Data Linkage Plans, if Applicable

For studies involving linkage of registry data with another data source, describe the other data source and its appropriateness and limitations for addressing specific hypotheses. Consider any additional requirements that may influence successful linkage, such as information needed to match patients, selection of data elements, and definitions used.

8.2.2 Plan Follow-up Based on the Registry Objective(s)

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

8.2.3 Describe Data Safety and Security

Research proposals and protocols should describe data use agreements, informed consent, data security, and approaches to protecting security including risk of re-identification of patients. If using previously collected data, describe how these address the risk of re-identification of patients and the actual use of data compared with the originally designed and consented use of the data.

8.2.4 Take Appropriate Steps to Ensure Data Quality

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

8.2.5 Document and Explain Any Modifications to the Protocol

Modifications to a registry protocol may be necessary for a variety of reasons. When modifications are necessary, document and explain any modifications to the formal study protocol.

8.2.6 Collect Data Consistently

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

8.2.7 Enroll and Follow Patients Systematically

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

8.2.8 Monitor and Take Actions to Keep Loss to Follow-up to an Acceptable Minimum

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

Standards for Data Registries (continued)

8.2.9 Use Appropriate Statistical Techniques to Address Confounding

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

Rationale for the Standards

Well-constructed and implemented registry studies promote patient-centeredness by providing timely data relevant to clinician and patient decision-making.

Registries are most likely to generate usable inferences if their construction is based on a protocol related to at least one question which includes plans for enrollment, patient follow-up, and data linkage. Such protocols must also include details of consent procedures and confidentiality protections that take into account the possibility of re-identification.

Standards for Studies of Diagnostic Tests (See *Standards in the Design, Conduct and Evaluation of Diagnostic Testing* http://www.pcori.org/what-we-do/methodology/)

Patients, caregivers, and providers need specific information about the expected benefits and harms of a diagnostic test in their particular circumstances. Diagnostic tests, however, are difficult to study, and even major studies of diagnostic tests often have serious flaws in design. The difficulty of conducting research on diagnostic tests has consequences. When the research on a test is flawed, physicians who obtain the test may under- or overestimate the likelihood that a patient has a disease. Some diagnostic tests also expose patients to harms, including inconvenience, radiation, complications of downstream invasive procedures, and psychological harms such as anxiety and stress.

A fundamental issue in diagnostic test research is how to define the benefit of a test. Tests generate information but do not directly produce a better outcome for the patient. Rather, to improve outcomes, the test result must be used effectively—for example, by helping with a decision about which treatment to use, what lifestyle changes might avert or ameliorate disease, or what additional tests should be done.

Ultimately, if tests benefit patients they have to improve health outcomes. The impact of diagnostic testing on patient outcomes has been traditionally understudied in clinical research. An understanding of the impact of a diagnostic test on patient outcomes is critical for patients and providers to understand how the diagnostic test may best be used in clinical care. Many studies of diagnostic tests are not designed to identify all of the pertinent effects of testing on patients, particularly long-term benefits and harms and cognitive, emotional, social, and behavioral effects.(48)

The effect of a diagnostic test may also depend on the choices clinicians and patients make and how well they carry out those choices. A challenge for investigators designing a study of a diagnostic test is whether to specify the actions clinicians should take based on test results (such as observation, further testing, or treatment) or to leave those responses to the discretion of patients and their providers. A related challenge is to understand and specify the pathways.

Additionally, a test must provide *new* information—information that is not already known or available. Failing to take account of information available from clinical features, including signs, symptoms, and other tests, is a common flaw in studies of diagnostic tests.

Diagnostic technology also tends to evolve rapidly. Although standards exist for the reporting of diagnostic or predictive accuracy studies (e.g., STARD), standards have not been established for studies of the impact of diagnostic tests on subsequent care or patients' outcomes.

Many studies of diagnostic tests evaluate how a test affects the likelihood of a disease—its accuracy—but do not collect data on what was done because of the test results, or how those actions turned out. Research on accuracy is essential, but it is not enough. For example, a test to see whether a life-threatening infection is caused by a particular type of bacteria might be perfectly accurate, but if it takes three weeks to get the result, it would be useless: all decisions about treating such a patient need to be made quickly. An accurate test could also fail to improve outcomes if the

information it provides is duplicative and does not change any decisions. Suppose, for example, that there is a new test that can help determine whether acute chest pain is due to heart disease or has a less serious cause. In actual practice, even if it is reasonably accurate, the use of this test might not be helpful if everyone who undergoes it must still undergo cardiac catheterization to be sure about the cause of the pain. In this situation the test is just an add-on—a test that provides information that will be obtained anyway whether the result is positive or negative. The use of the test could even be harmful if, for patients who are having a heart attack, the test increases the time until the patient can get a cardiac catheterization and definitive treatment.

Considerable variation exists in how well studies of diagnostic tests have defined and described their methods. Diagnostic tests are studied through both experiments (including randomized controlled trials) and observational studies (including reviews of medical records and registries). Trials can be complex and take years to complete. A wide variety of observational designs is used to assess the accuracy and impact of diagnostic tests. Adapting the principles of randomized trial design can be a powerful strategy for designing observational studies of diagnostic tests.(49) Guidance for reporting studies of diagnostic accuracy relies on these principles.(50-53) The U.S. Food and Drug Administration(54) and the CONSORT group also offer guidance about the evaluation of diagnostic tests.

Standards for Studies of Diagnostic Tests

8.3.1 Specify Clinical Context and Key Elements of Diagnostic Test Study Design

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

8.3.2 Study Design Should Be Informed by Investigations of the Clinical Context of Testing

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

8.3.3 Assess the Effect of Factors Known to Affect Diagnostic Performance and Outcomes

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

8.3.4 Structured Reporting of Diagnostic Comparative Effectiveness Study Results

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

8.3.5 Give Preference to Randomized Designs of Studies of Test Outcomes

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

Scope of the Standards

In addition to diagnosis, medical tests have many other uses: to predict the risk of developing a disease; to establish the prognosis of a disease or condition; to predict the chance of a response or of serious adverse effects to a treatment; and, especially for imaging studies, to identify the anatomic location and extent of disease. These uses will be addressed in future work by the Methodology Committee. Standards for systematic reviews of test accuracy are being developed by other groups.(55-60) In future work, the Committee will also address the use of decision analysis and simulation modeling as related to the use of diagnostic tests.(61)

Rationale for the Standards

The standards included in this report concern tests that are done to diagnosis a disease that, ultimately, can be confirmed by a more definitive test or procedure.

The standards reflect the principles that accuracy alone is often not a sufficient definition or measure of the benefit of a test, that a test or testing strategy should be compared with alternatives, and that large prospective studies are often the best way to conduct these comparisons.(62)

The fact that diagnostic tests mostly generate information for decision-making rather than outcomes and that diagnostic testing procedures can change rapidly requires that study designers pay particular attention to fully specify and understand the test(s) to be studied, the use of the test(s), disease outcomes of interest, existing practices, and the efficacy/effectiveness of interventions. Precise specification of the design and rationale of the study evaluating a diagnostic test clarifies to which patients in what settings the results apply and supports the overall scientific validity and transparency of the study to providers and stakeholders. Prospective randomized trials can minimize problems of selection bias and confounding by indication, improving the validity of results. Randomized trials of diagnostic tests are infrequent due to cost and complexity, though one well-designed multi-site trial can yield definitive answers and thus may be more cost-effective than multiple observational studies. As patients enrolled in RCT tend to be more homogenous, observational studies may be used to augment results from the definitive prospective studies.

Recommended Actions

In order to support the design-specific standards covered in this chapter, PCORI should:

- Develop additional guidance specific to adaptive trials.
- Support development and use of software for adaptive trials that can simulate complex designs.
- Broaden experience with adaptive trials for PCOR, perhaps through funding of a cohort of adaptive trials on priority topic areas.
- Mentor investigators and develop a "how to" guide and a forum to share experiences with adaptive trials.
- Develop coursework and other training opportunities for statisticians and other methodologists interested in developing expertise in adaptive trials.
- Sponsor an Institute of Medicine committee to develop standards for research on medical tests.

Research Recommendations

For registry studies, PCORI should:

- Develop analytic techniques for addressing measured and unmeasured confounding.
- Develop analytic techniques for handling missing data that can be used in registry studies.
- Develop improved strategies for linking data while maintaining privacy protections and assuring that link data do not lead to re-identification in de-identified data.
- Develop innovative ways to reduce loss to follow-up as registries encompass longer time periods.

Chapter 9: Next Steps

This report is first submitted to the PCORI Board of Governors, and is expected to be publicly posted immediately for information. Following review by the Board and appropriate revision, the report will be formally posted for public comment during the summer of 2012, with expected response to public comment, revisions, and final approval in the fall of 2012. This chapter briefly summarizes the Committee's current thinking about steps beyond this first report. In addition to the core tasks named in the authorizing legislation, the Committee also notes here several topics related to its mission (not otherwise discussed in the report) that are of ongoing interest.

Developing Standards

As noted in Chapter 1, the set of methodological standards proposed here for patient-centered outcomes research is an important milestone, not the destination. The legislation establishing PCORI and its Methodology Committee directs that these standards be periodically updated (Appendix E-5). The Committee expects to reconsider, refine, and widen the scope of the standards to include the full spectrum of PCOR questions and approaches. The Committee also expects to 1) refine the methods used to develop the standards, including improvements to the methods we used to identify, evaluate, and synthesize existing standards; 2) refine the methods for developing new standards in areas where existing standards are absent; and 3) review the empirical evidence supporting existing and new standards and evaluate their usefulness in specifying appropriate research designs and methods. As a core function of the Committee, further development of the standards is a prominent part of the "blueprint" for future work (see below).

Below is a partial listing of specific actions that the Committee intends to take:

1. Expand the inventory of research methods relevant to patient-centered CER for which standards are needed.

2. Distinguish between standards that are minimum requirements and those that may be aspirational or best practice but are not required.

3. Clarify standards addressing the selection of specific research design methods versus those whose focus is on other parts of the research enterprise (e.g., administrative or organizational).

4. Specify and support new research to strengthen methods relevant to CER and PCOR.

5. Further develop the criteria that the MC will use in recommending standards.

6. Refine processes to use Methodology Committee members, PCORI scientific staff, external groups (e.g., Institute of Medicine, AHRQ, or professional societies), consultants, and other stakeholders in locating, assessing, and synthesizing existing standards and developing new standards.

The Translation Table

The translation table discussed within this report (Chapter 6) will be expanded to include more examples, methodological issues, and approaches. With input and advice from the broad stakeholder community, the Committee needs to decide on the list of research categories that the suite of translation tables will eventually cover and further expects that each will require the development of translation tools appropriate to the various stakeholder groups. An important goal will be to validate the framework and evaluate its usefulness.

Supporting Adoption of Methodological Standards

The Methodology Committee and PCORI expect to develop a comprehensive, coordinated approach to promote the uptake of PCORI methods standards. This includes engaging all stakeholders who might use the standards, collaborating with existing entities and initiatives to strengthen research practices and facilitate use of the standards, creating reporting and surveillance mechanisms, and over time considering the value of enforcement functions. Future activities might include the development of training resources, checklists, and other tools to support researchers' decisions and practices, as well as checklists and other decision support tools for peer review and others. The Committee assumes that PCORI itself will apply these standards across all phases of the research enterprise, and, at a minimum, the Committee expects to specify how adherence to the standards can be measured.

Patient-Centeredness

As noted in Chapter 4, the Committee intends to focus on encouraging research that will develop evidence to fill three important gaps in patient/stakeholder engagement knowledge: 1) the

consequences of patient engagement in research on health decisions and clinical outcomes; 2) the specific consequences of patient engagement on the research process; and 3) the patient engagement methods that are most effective and for which specific populations.

Research Priorities

The Committee expects further development of the major approaches for research prioritization presented in Chapter 5: 1) methods for topic generation, 2) gap analysis using systematic reviews, 3) value of information (VOI) analysis, and 4) peer and stakeholder review. The current status of the Committee's thinking on this topic, including plans for further work in each of the four domains, is fully described in the chapter and is not repeated here. The Committee understands the challenges it faces in evaluating the recommended actions to support its proposed standards. The Committee aims to further develop the standards and integrate them into a robust process that can address the needs of PCOR.

Building Infrastructure: The Use of Electronic Medical Records

The adoption of electronic medical records (EMRs) represents a multi-billion dollar national infrastructure investment on the part of U.S. healthcare delivery and payment organizations. The potential of EMR to answer significant PCOR questions is an issue of great interest to the Methodology Committee. The Committee hopes to eventually make recommendations regarding the use of EMRs in PCOR research, but first aims to better understand the medical, technical, political, and financial pressures that have as yet prevented widespread use of EMRs in conducting high-quality PCOR research.

The Committee asked Deloitte Consulting to conduct a series of interviews to understand key initiatives, gaps, challenges, and potential roles PCORI can play in this environment. Deloitte conducted 44 interviews including representatives from government, academia, commercial organizations, providers, and associations. A sample of interview questions, a summary of responses, and the full responses are available from PCORI.

The Committee has scheduled a workshop July 2-3, 2012, at Stanford University (including Webcast) with 60-80 participants from academia, professional and patient associations, commercial, government, and state organizations, and rural and underserved communities. Some of the expected

outcomes/goals of this workshop are to guide PCORI's role and identify areas in which it can make a difference in health information technology as it relates to CER/PCOR; to develop relationships in the informatics, methodology, and patient communities; and to further understand the current landscape and where investments are, will, and can be made. The Committee's next steps on this topic will be decided after the conference, based on the original interviews, conference proceedings, and Committee deliberations.

Enhancing Capacity among the Research Community to Understand, Deploy, and Improve PCOR Methods

In partnership with AHRQ, NIH, and others, PCORI will develop plans to enhance understanding and appropriate use of methods through selected educational and training opportunities.

The Methodology Committee "Blueprint"

The Committee has been working on a blueprint for its future, clarifying its vision and mission and specifying a rough timeline for further work on standards, the translation table, dissemination, implementation, and evaluation extending through January 2014. A draft of the blueprint was submitted to PCORI staff in the fall of 2011 for consideration in the strategic planning process. The Committee views the blueprint as an internal working document subject to continual review and revision. Although we have already engaged the Board in providing feedback on earlier drafts, further comments are welcome.

References

1. Helfand M, Tunis S, Whitlock EP, Pauker SG, Basu A, Chilingerian J, et al. A CTSA agenda to advance methods for comparative effectiveness research. Clin Transl Sci. 2011 Jun;4(3):188-98.

2. Lohr KN. Emerging methods in comparative effectiveness and safety: symposium overview and summary. Med Care. 2007 Oct;45(10 Supl 2):S5-8.

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

4. Kievit W, Fransen J, Oerlemans AJ, Kuper HH, van der Laar MA, de Rooij DJ, et al. The efficacy of anti-TNF in rheumatoid arthritis, a comparison between randomised controlled trials and clinical practice. Ann Rheum Dis. 2007 Nov;66(11):1473-8.

5. Sackett DL. Evidence-based medicine: How to practice and teach EBM (2nd ed.). Edinburgh; New York: Churchill Livingstone; 2000.

6. Krumholz H. Missing Data: The Elephant That's Not in the Room. 2012; Available from: http://www.cardioexchange.org/voices/missing-data-the-elephant-thats-not-in-the-room/.

7. Ransohoff DF. How to improve reliability and efficiency of research about molecular markers: roles of phases, guidelines, and study design. J Clin Epidemiol. 2007 Dec;60(12):1205-19.

8. Chobanian AV. Does it matter how hypertension is controlled? N Engl J Med. 2008 Dec 4;359(23):2485-8.

9. Wassertheil-Smoller S, Psaty B, Greenland P, Oberman A, Kotchen T, Mouton C, et al. Association between cardiovascular outcomes and antihypertensive drug treatment in older women. JAMA. 2004 Dec 15;292(23):2849-59.

10. Boger-Megiddo I, Heckbert SR, Weiss NS, McKnight B, Furberg CD, Wiggins KL, et al. Myocardial infarction and stroke associated with diuretic based two drug antihypertensive regimens: population based case-control study. BMJ. 2010;340:c103.

11. Jamerson K, Weber MA, Bakris GL, Dahlof B, Pitt B, Shi V, et al. Benazepril plus amlodipine or hydrochlorothiazide for hypertension in high-risk patients. N Engl J Med. 2008 Dec 4;359(23):2417-28.

12. O'Riordan M. Letters to New England Journal of Medicine question details about ACCOMPLISH findings. March 11, 2009; Available from: http://www.theheart.org/article/947523.doc.

13. National Research Council. Find What Works in Health Care: Standards for Systematic Reviews. Washington, DC: The National Academies Press; 2011.

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

15. Institute of Medicine (IOM). Consensus Report. Knowing What Works in Health Care: A Roadmap for the Nation. 2008 [cited 2012 January 10]; Available from: http://www.iom.edu/Reports/2008/Knowing-What-Works-in-Health-Care-A-Roadmap-for-the-Nation.aspx.

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

17. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. JAMA. 2003 Sep 24;290(12):1624-32.

18. Helfand M, Aguilar-Gaxiola SA, Selker HP. Comparative effectiveness research: an approach to avoiding "overgeneralized medicine". Clin Transl Sci. 2009 Dec;2(6):444-5.

19. Luce BR, Kramer JM, Goodman SN, Connor JT, Tunis S, Whicher D, et al. Rethinking randomized clinical trials for comparative effectiveness research: the need for transformational change. Ann Intern Med. 2009 Aug 4;151(3):206-9.

20. Guyatt G, Rennie D, Meade M, Cook D. Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice, 2nd Edition. United States: The McGraw-Hill Companies, Inc; 2008.

21. JAMA evidence Glossary (Derived from: Users' Guides to the Medical Literature: A Manual for Evidence-Based Practice, 2nd Edition, The Rational Clinical Examination: Evidence-Based Clinical Diagnosis and Care at the Close of Life: Evidence and Experience. Updated February 2012.) See Patient-important outcomes. Available from: http://jamaevidence.com/glossary/P.

22. Viswanathan M, Ammerman A, Eng E, Garlehner G, Lohr K, Griffith D, et al. Community-Based Participatory Research: Assessing the Evidence: Summary. AHRQ Evidence Report Summaries. Rockville (MD): Agency for Healthcare Research and Quality (US); 2004 Aug.

23. Deverka PA, Schully SD, Ishibe N, Carlson JJ, Freedman A, Goddard KA, et al. Stakeholder assessment of the evidence for cancer genomic tests: insights from three case studies. Genet Med. 2012 Apr 5.

24. 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 Oct;27(4):391-9.

25. Gagnon MP, Desmartis M, Lepage-Savary D, Gagnon J, St-Pierre M, Rhainds M, 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 Jan;27(1):31-42.

26. Agency for Healthcare Research and Quality (AHRQ). Effective Health Care Program. Available from: http://effectivehealthcare.ahrq.gov.

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

28. Oliver SR, Rees RW, Clarke-Jones L, Milne R, Oakley AR, Gabbay J, et al. A multidimensional conceptual framework for analysing public involvement in health services research. Health Expect. 2008 Mar;11(1):72-84.

29. Patient Partner. The active involvement of Patients and Patient organisations in (the) clinical trial (development process). 2012; Available from: http://www.patientpartner-europe.eu/en/resources/active-involvement-in-the-process.

30. Carey T, Sanders G, Viswanathan M, Trikalinos T, Kato E, Chang S. Framework for Considering Study Designs for Future Research Needs. Methods Future Research Needs Paper No. 8 (Prepared by the RTI-UNC Evidence-based Practice Center under Contract No. 290-2007-10056-I.) AHRQ Publication No. 12-EHC048-EF. Rockville, MD: Agency for Healthcare Research and Quality. March 2012.

31. McKoy J, Hartmann K, Jerome R. Future Research Needs for Outcomes of Weight Gain in Pregnancy. Rockville, MD: Agency for Healthcare Research and Quality, 2010. AHRQ Publication No. 11-EHC004-EF.

32. Raiffa H, Schlaifer R. Applied Statistical Decision Theory. Harvard Business School, Colonial Press. 1961.

33. Meltzer D. Addressing uncertainty in medical cost-effectiveness analysis implications of expected utility maximization for methods to perform sensitivity analysis and the use of cost-effectiveness analysis to set priorities for medical research. J Health Econ. 2001 Jan;20(1):109-29.

34. Bravata DM, McDonald KM, Shojania KG, Sundaram V, Owens DK. Challenges in systematic reviews: synthesis of topics related to the delivery, organization, and financing of health care. Ann Intern Med. 2005 Jun 21;142(12 Pt 2):1056-65.

35. Buede D. Influence diagrams: A practitioner's perspective. Decision Analysis. 2005;2(4):235-37.

36. Hernan MA, Alonso A, Logan R, Grodstein F, Michels KB, Willett WC, et al. Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease. Epidemiology. 2008 Nov;19(6):766-79.

37. Goldfine AB, Kaul S, Hiatt WR. Fibrates in the treatment of dyslipidemias--time for a reassessment. N Engl J Med. 2011 Aug 11;365(6):481-4.

38. Lagakos SW. The challenge of subgroup analyses--reporting without distorting. N Engl J Med. 2006 Apr 20;354(16):1667-9.

39. 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.

40. Arnold K. Journal to encourage analysis by sex/ethnicity. J Natl Cancer Inst. 2000 Oct 4;92(19):1561.

41. Food and Drug Administration. Draft Guidance for Industry and Food and Drug Administration Staff: Evaluation of Sex Differences in Medical Device Clinical Studies. Rockville, MD: U.S. Department of Health and Human Services. December 19, 2011.

42. Galvagno SM, Jr., Haut ER, Zafar SN, Millin MG, Efron DT, Koenig GJ, Jr., et al. Association between helicopter vs ground emergency medical services and survival for adults with major trauma. JAMA. 2012 Apr 18;307(15):1602-10.

43. National Research Council. The Prevention and Treatment of Missing Data in Clinical Trials. Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press. 2010; Available from: http://www.nap.edu/openbook.php?record_id=12955.

44. Diamond GA, Bax L, Kaul S. Uncertain effects of rosiglitazone on the risk for myocardial infarction and cardiovascular death. Ann Intern Med. 2007 Oct 16;147(8):578-81.

45. OECD. OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data. Available from: http://www.oecd.org/document/18/0,3746,en_2649_34223_1815186_1_1_1_1,00.html.

46. Fiore LD, Brophy M, Ferguson RE, D'Avolio L, Hermos JA, Lew RA, et al. A point-of-care clinical trial comparing insulin administered using a sliding scale versus a weight-based regimen. Clin Trials. 2011 Apr;8(2):183-95.

47. Muss HB, Berry DA, Cirrincione CT, Theodoulou M, Mauer AM, Kornblith AB, et al. Adjuvant chemotherapy in older women with early-stage breast cancer. N Engl J Med. 2009 May 14;360(20):2055-65.

48. Bossuyt PM, McCaffery K. Additional patient outcomes and pathways in evaluations of testing. Med Decis Making. 2009 Sep-Oct;29(5):E30-8.

49. Lord SJ, Irwig L, Bossuyt PMM. Using the Principles of Randomized Controlled Trial Design To Guide Test Evaluation. Medical Tests-White Paper Series. Rockville (MD): Agency for Healthcare Research and Quality (US). 2009.
50. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. BMJ. 2003 Jan 4;326(7379):41-4.
51. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, et al. The STARD statement for reporting studies of diagnostic accuracy: explanation and elaboration. Clin Chem. 2003 Jan;49(1):7-18.
52. Whiting PF, Rutjes AW, Westwood ME, Mallett S, Deeks JJ, Reitsma JB, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. Ann Intern Med. 2011 Oct 18;155(8):529-36.
53. Whiting PF, Weswood ME, Rutjes AW, Reitsma JB, Bossuyt PN, Kleijnen J. Evaluation of QUADAS, a tool for the quality assessment of diagnostic accuracy studies. BMC Med Res Methodol. 2006;6:9.
54. U.S. Food and Drug Administration. Statistical Guidance on Reporting Results from Studies Evaluating Diagnostic Tests. 2007 [updated March 13, 2007]; Available from: http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071148.htm.
55. Reitsma J, Rutjes A, Whiting P, Vlassov V, Leeflang M, Deeks J. Chapter 9: Assessing methodological quality. In: Deeks JJ BP, Gatsonis C, editor. Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy Version 100: The Cochrane Collaboration; 2009.
56. Matchar D. Introduction to the methods guide for medical test reviews. J Gen Intern Med. 2011.
57. Santaguida P, Riley C, Matchar D. Chapter 5: Assessing risk of bias as a dojmain of quality in medical test studies. J Gen Intern Med. 2012.
58. Trikalinos T, Balion C, Colemlan C. Chapter 8: meta-analysis of test performance when there is a 'Gold Standard'. J Gen Intern Med. 2012a.
59. Trikalinos T, Ballion C. Chapter 9: Options for summarizing medical test performance in the absence of a 'Gold Standard'. J Gen Intern Med. 2012b.
60. Trikalinos T, Kulasingam S, Lawrence W. Chapter 10: Deciding Whether to Complement a Systematic Review of Medical Tests with Decision Modeling. J Gen Intern Med. 2012c.
61. Trikalinos TA, Siebert U, Lau J. Decision-analytic modeling to evaluate benefits and harms of medical tests: uses and limitations. Med Decis Making. 2009 Sep-Oct;29(5):E22-9.
62. 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.

Appendix A: Standards for Patient-Centered Outcomes Research

Standards by Title and Number

Standard Type	Number and Title
Standards for Formulating Research Questions	3.1.1 Develop a Formal Study Protocol 3.1.2 Identify Specific Populations and Health Decision(s) Affected by the Research 3.1.3 Identify and Assess Participant Subgroups 3.1.4 Select Appropriate Interventions and Comparators 3.1.5 Measure Outcomes that People in the Population of Interest Notice and Care About
Patient-Centeredness	4.1.1 Engage Patient Informants, Persons Representative of the Population of Interest, in All Phases of Patient-Centered Outcomes Research (PCOR) 4.1.2 Identify, Select, Recruit, and Retain Study Participants Representative of the Spectrum of the Population of Interest Facing the Health Decision of Interest and Ensure that Data Are Collected Thoroughly and Systematically from All Study Participants 4.1.3 Use Patient-Reported Outcomes (PROs) When Patients or People at Risk of a Condition Are the Best Source of Information 4.1.4 Develop and Implement a Dissemination Assessment to Achieve Broad Awareness of Study Results
Research Prioritization	5.1.1 Use Systematic Reviews to Identify Gaps in Evidence 5.1.2 Protect Independence in Peer Review of Research Funding Proposals 5.1.3 Peer Review of Research Funding Proposals for Problems Affecting Minorities and Disadvantaged Segments of the Population
General and Crosscutting Methods for All PCOR	7.1.1 Assess Data Source Adequacy 7.1.2 A priori, Specify Plans for Data Analysis that Correspond to Major Aims 7.1.3 Document Validated Scales and Tests 7.1.4 Use Sensitivity Analyses to Determine the Impact of Key Assumptions 7.1.5 Provide Sufficient Information in Reports to Allow for Assessments of the Study's Internal and External Validity
Causal Inference	7.2.1 Define Analysis Population Using Information Available at Study Entry 7.2.2 Describe Population that Gave Rise to the Effect Estimate(s) 7.2.3 Precisely Define the Timing of the Outcome Assessment Relative to the Initiation and Duration of Intervention 7.2.4 Measure Confounders before Start of Exposure 7.2.5 Assess Propensity Score Balance 7.2.6 Assess Instrumental Variable Assumptions

Standards by Title and Number (continued)

Standard Type	Number and Title
Heterogeneity Of Treatment Effects (HTE)	7.3.1 State the Goals of HTE Analyses 7.3.2 For Confirmatory and Descriptive HTE Analyses, Pre-specify Subgroups and Outcomes; for Confirmatory HTE Analyses, Pre-specify Hypotheses for Each Subgroup Effect 7.3.3 For Confirmatory HTE Analyses, Report a priori Statistical Power 7.3.4 For Any HTE Analysis, Perform an Interaction Test and Report Sufficient Information on Treatment Effect Estimates 7.3.5 For Exploratory HTE Analyses, Discuss Findings in the Context of Study Design and Prior Evidence 7.3.6 For Any HTE Analysis, Report All Pre-specified Analyses and, at Minimum, the Number of Post-hoc Analyses, Including Number of Subgroups and Outcomes Analyzed
Missing Data	7.4.1 Describe in Protocol Methods to Prevent and Monitor Missing Data 7.4.2 Describe Statistical Methods to Handle Missing Data in Protocol 7.4.3 Use Validated Methods to Deal with Missing Data that Properly Account for Statistical Uncertainty Due to Missingness, Such as Multiple Imputation—All Forms of Single Imputation Are Discouraged 7.4.4 Record and Report All Reasons for Dropout and Missing Data, and Account for All Patients in Reports 7.4.5 Examine Sensitivity of Inferences to Missing Data Methods and Assumptions, and Incorporate into Interpretation
Data Networks	7.5.1 Data Integration Strategy 7.5.2 Risk Assessment Strategy. 7.5.3 Identity Management and Authentication of Individual Researchers 7.5.4 Intellectual Property Policies 7.5.5 Standardized Terminology Encoding of Data Content 7.5.6 Metadata Annotation of Data Content 7.5.7 Common Data Model
Adaptive Trials	8.1.1 Specify Planned Adaptations and Primary Analysis 8.1.2 Evaluate Statistical Properties of Adaptive Design 8.1.3 Specify Structure and Analysis Plan for Bayesian Adaptive Randomized Clinical Trial Designs 8.1.4 Ensure Clinical Trial Infrastructure Is Adequate to Support Planned Adaptation(s) 8.1.5 Use the CONSORT Statement, with Modifications, to Report Adaptive Randomized Clinical Trials

Standards by Title and Number (continued)

Standard Type	Number and Title
Data Registries	8.2.1 Describe Data Linkage Plans, if Applicable
	8.2.2 Plan Follow-up Based on the Registry Objective(s)
	8.2.3 Describe Data Safety and Security
	8.2.4 Take Appropriate Steps to Ensure Data Quality
	8.2.5 Document and Explain Any Modifications to the Protocol
	8.2.6 Collect Data Consistently
	8.2.7 Enroll and Follow Patients Systematically
	8.2.8 Monitor and Take Actions to Keep Loss to Follow-up to an Acceptable Minimum
	8.2.9 Use Appropriate Statistical Techniques to Address Confounding
Diagnostic Tests	8.3.1 Specify Clinical Context and Key Elements of Diagnostic Test Study Design
	8.3.2 Study Design Should Be Informed by Investigations of the Clinical Context of Testing
	8.3.3 Assess the Effect of Factors Known to Affect Diagnostic Performance and Outcomes
	8.3.4 Structured Reporting of Diagnostic Comparative Effectiveness Study Results
	8.3.5 Give Preference to Randomized Designs of Studies of Test Outcomes

Full Text of Standards

	Standards for Formulating Research Questions
3.1.1	**Develop a Formal Study Protocol** In addition to the general study protocol standards, registry studies should include a formal study protocol specifying: at least one purpose of the registry (e.g., effectiveness, safety, natural history of disease, quality improvement, or other); data sources and linkage plans, if any; measure(s) of effect; and use of any standardized data dictionaries (nationally or internationally accepted).
3.1.2	**Identify Specific Populations and Health Decision(s) Affected by the Research** To produce information that is meaningful and useful to people when making specific health decisions, research proposals and protocols should describe: 1) the specific health decision the research is intended to inform; 2) the specific population for whom the health decision is pertinent; and 3) how study results will inform the health decision.
3.1.3	**Identify and Assess Participant Subgroups** In designing studies, researchers should identify participant subgroups of interest and, where feasible, design the study with adequate precision and power to reach conclusions specific to these subgroups. In addition, subgroup information should be reported for later systematic reviews.
3.1.4	**Select Appropriate Interventions and Comparators** When evaluating an intervention, the comparator treatment(s) must be chosen to enable accurate evaluation of effectiveness or safety compared to other viable options for similar patients. Researchers should make explicit what the comparators are and how they were selected, focusing on clearly describing how the chosen comparator(s) define the causal question, reduce the potential for biases, and allow direct comparisons. Generally, non-use (or no specific treatment) comparator groups should be avoided unless no specific treatment is a likely option in standard care.
3.1.5	**Measure Outcomes that People in the Population of Interest Notice and Care About** Identify and select outcomes the population of interest notices and cares about (e.g., survival, function, symptoms, health-related quality of life) and that inform an identified health decision. Define outcomes clearly, especially for complex conditions or outcomes that may not have established clinical criteria. Provide information that supports the selection of outcomes as meeting the criteria of "clinically meaningful," "patient-centered," and "relevant to decision-makers," such as patient and decision-maker input from meetings or surveys or published literature relevant to the question of interest. Select outcomes based on input directly elicited from patient informants, persons representative of the population of interest, either in previous studies or in the proposed research.

Standards Associated with Patient-Centeredness

4.1.1 Engage Patient Informants, Persons Representative of the Population of Interest, in All Phases of Patient-Centered Outcomes Research (PCOR)

Research proposals should 1) describe how patient informants will be: identified, recruited, and retained; involved in determining the study design and monitoring of its conduct; and involved in dissemination of research results, and 2) state how the research process will follow PCOR principles of trust, transparency, co-learning, respect, and partnership. Patient informants include individuals who have the condition or who are at risk of the condition, and, as relevant, their surrogates or caregivers. At a minimum, patient informants should be engaged in formulating research questions; defining essential characteristics of study participants, comparators, and outcomes; monitoring study conduct and progress; and disseminating results.

4.1.2 Identify, Select, Recruit, and Retain Study Participants Representative of the Spectrum of the Population of Interest Facing the Health Decision of Interest and Ensure that Data Are Collected Thoroughly and Systematically from All Study Participants

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

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

4.1.3 Use Patient-Reported Outcomes When Patients or People at Risk of a Condition Are the Best Source of Information

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

4.1.4 Develop and Implement a Dissemination Assessment to Achieve Broad Awareness of Study Results

PCOR research proposals and protocols must include an assessment that describes how the project and the composition of the research team supports dissemination and the anticipated facilitators, barriers and potential strategies for dissemination to key stakeholder groups, including patients and individuals at risk of a condition, clinicians and other health care system staff, and policy leaders. Effective dissemination includes the reporting of results in a manner understandable to each target audience, information regarding the relevance of the results for decision-making (recognizing that research findings from a single study alone should not necessarily affect decision-making or practice), along with attention to how the results can be incorporated into health decision-making if applicable. The plan must specify how the dissemination strategy is expected to affect the identified health decisions and how dissemination engages the study participants or the population of interest. Requiring research dissemination, as well as engagement of patients and other stakeholders at this stage of research, represents a cultural shift for many institutions and researchers.

Standards for Prioritizing Research

5.1.1 Use Systematic Reviews to Identify Gaps in Evidence

Gap analysis of systematic reviews should be used as part of the process of identifying and prioritizing research gaps to establish funding priorities by PCORI.

5.1.2 Protect Independence in Peer Review of Research Funding Proposals

Adopted methods of peer review should aim to safeguard independence between reviewers and those being reviewed.

5.1.3 Ensure Adequate Representation of Minorities and Disadvantaged Segments of the Population in Peer Review of Research Funding Proposals

Approaches to topic generation in PCOR should involve both consultative and collaborative functions.

7.1.1 Assess Data Source Adequacy

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

7.1.2 A priori, Specify Plans for Data Analysis that Correspond to Major Aims

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

7.1.3 Document Validated Scales and Tests

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

7.1.4 Use Sensitivity Analyses to Determine the Impact of Key Assumptions

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

7.1.5 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 has brought together all reporting guidelines that have been developed using formal approaches, many of which have been adopted by journals, such as CONSORT (for randomized clinical trials), STARD (for diagnostic tests), and STROBE (for observational studies).

Causal Inference Standards

7.2.1 Define Analysis Population Using Information Available at Study Entry

Decisions about whether patients are included in an analysis should be based on information available at each patient's time of study entry and not based on future information, such as future changes in exposure.

7.2.2 Describe Population that Gave Rise to the Effect Estimate(s)

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

7.2.3 Precisely Define the Timing of the Outcome Assessment Relative to the Initiation and Duration of Intervention

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

7.2.4 Measure Confounders before Start of Exposure

In general, variables for use in confounding adjustment (either in the design or analysis) should be ascertained and measured prior to the first exposure to the therapy (or therapies) under study.

7.2.5 Assess Propensity Score Balance

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

7.2.6 Assess Instrumental Variable Assumptions

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

Standards for Heterogeneity of Treatment Effect (HTE)

7.3.1 State the Goals of HTE Analyses

State the inferential goal of each HTE analysis; identify each analysis as confirmatory, descriptive, or exploratory. See Table 7.1 comparing the different types of HTE analyses.

7.3.2 For Confirmatory and Descriptive HTE Analyses, Pre-specify Subgroups and Outcomes; for Confirmatory HTE Analyses, Pre-specify Hypotheses for Each Subgroup Effect

The study protocol should unambiguously pre-specify planned confirmatory HTE analyses. Pre-specification of confirmatory HTE analyses should include a public record with a clear statement of the hypotheses the study will evaluate, including the definitions of subgroup variables and outcomes, and the direction of the expected treatment effects. Prior evidence should be available for review and the study protocol should present this evidence clearly. The hypotheses for descriptive HTE need not be pre-specified; rather, specify the subgroups to be studied, as one goal is to facilitate future meta-analyses.

7.3.3 For Confirmatory HTE Analyses, Report a priori Statistical Power

Studies should calculate and report the power to detect treatment effects in each subgroup and to detect the interaction between the treatment and the subgrouping variable (i.e., the power to test whether the effects are statistically different between particular subgroups).

7.3.4 For Any HTE Analysis, Perform an Interaction Test and Report Sufficient Information on Treatment Effect Estimates

To detect differences in treatment effect between subgroups, use an interaction test (i.e., test whether the interaction between the treatment indicator and the subgroup variable is statistically significant).

Within each subgroup level, studies should present treatment effect estimates, standard errors, and 95 percent confidence intervals. Studies should also report the P-value for the interaction test for each subgrouping variable. For descriptive analyses, studies should also consider presenting a forest plot as a visual summary of the results, although such forest plots should not be used to infer HTE.

7.3.5 For Exploratory HTE Analyses, Discuss Findings in the Context of Study Design and Prior Evidence

Exploratory HTE analyses should be presented in the context of whether they are consistent with prior evidence and how well the study design addresses the HTE question. These considerations are more important than P-values for inferences.

7.3.6 For Any HTE Analysis, Report All Pre-specified Analyses and, at Minimum, the Number of Post-hoc Analyses, Including Number of Subgroups and Outcomes Analyzed

Studies must report the results of all the HTE analyses that were pre-specified in the study protocol or grant application, regardless of their statistical significance. Reports of exploratory HTE analyses that did not pre-specify subgroups should clearly report the number of subgroups and outcomes analyzed.

Standards for Preventing and Handling Missing Data

7.4.1 Describe in Protocol Methods to Prevent and Monitor Missing Data

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

7.4.2 Describe Statistical Methods to Handle Missing Data in Protocol

Statistical methods for handling missing data should be pre-specified in study protocols, and their associated assumptions stated in a way that can be understood by stakeholders. The reasons for missing data should be considered in the analysis. This standard applies to all study designs for any type of research question.

7.4.3 Use Validated Methods to Deal with Missing Data that Properly Account for Statistical Uncertainty Due to Missingness, Such as Multiple Imputation; All Forms of Single Imputation Are Discouraged

Statistical inference of intervention effects or measures of association should account for statistical uncertainty attributable to missing data. This means that methods used for imputing missing data should have valid type I error rates and that confidence intervals should have the nominal coverage properties. This standard applies to all study designs for any type of research question. Multiple imputation methods satisfy this condition, along with various likelihood-based and other validated methods. Single imputation methods like last observation carried forward and baseline observation carried forward are discouraged as the primary approach for handling missing data in the analysis.

7.4.4 Record and Report All Reasons for Dropout and Missing Data, and Account for All Patients in Reports

Whenever a participant discontinues some or all types of participation in a research study, the investigator should document the following: 1) the reason for discontinuation; 2) who decided that the participant would discontinue; and 3) whether the discontinuation involves some or all types of participation. Investigators should continue to collect information on key outcomes for participants who discontinue their protocol-specified intervention. This standard applies to all prospective study designs that aim to assess intervention effectiveness.

All participants who enter the study should be accounted for in the report, whether or not they are included in the analysis. Describe and justify any planned reasons for excluding participants from analysis. This standard applies to all study designs for any type of research question.

7.4.5 Examine Sensitivity of Inferences to Missing Data Methods and Assumptions, and Incorporate into Interpretation

Examining sensitivity to the assumptions about the missing data mechanism (i.e., sensitivity analysis) should be a mandatory component of the study protocol, analysis, and reporting. This standard applies to all study designs for any type of research question.

Tables should be used to describe missing data studies. Potential bias resulting from imperfect definitions should be discussed with an estimate of the change in the direction and magnitude of the effect due to bias. These quantitative results should be incorporated into the interpretation of the study and reflected in the discussion section and possibly the abstract. If there are big effects, help the user further understand the reason for the missing data and the effect of the missingness on the findings.

Standards for Data Networks

7.5.1 Data Integration Strategy

In order for equivalent data elements from different sources to be harmonized (treated as equivalent), processes should be created and documented that either a) transform and standardize data elements prior to analysis or b) make transformation logic available that can be executed when data are extracted. The selected approach should be based on an understanding of the research domain of interest.

7.5.2 Risk Assessment Strategy

Data custodians should assess the uniqueness of records (i.e., no other records have the same values) of patient records to measure re-identification risk of data, and apply algorithms to ensure that the desired level of confidentiality is achieved to meet the need of the particular PCOR application.

7.5.3 Identity Management and Authentication of Individual Researchers

Develop a reliable process for verifying credentials of researchers who are granted access to a distributed research network and authenticating them.

7.5.4 Intellectual Property Policies

A research network should develop policies for the handling and dissemination of intellectual property (IP); networks should also have an ongoing process for reviewing and refreshing those policies. IP can include data, research databases, papers, reports, patents, and/or products resulting from research using the network. Guidelines should balance (1) minimizing impediments to innovation in research processes and (2) making the fruits of research widely accessible, particularly to the people who need them the most.

7.5.5 Standardized Terminology Encoding of Data Content

The data contents should be represented with standardized terminology systems to ensure that their meaning is unambiguously and consistently understood by the party using the data.

7.5.6 Metadata Annotation of Data Content

Semantic and administrative aspects of data contents should be annotated with a set of metadata items. Metadata annotation helps to correctly identify the intended meaning of a data element and facilitates an automated compatibility check among data elements. A data element is the entity and its property described with a given data. A data element is defined with a set of metadata.

7.5.7 Common Data Model

Individual data items should be assembled into a contextual environment that shows close or distant association among data. A common data model (CDM) specifies necessary data items that need to be collected and shared across participating institutes, clearly represents these associations and relationships among data elements, and promotes correct interpretation of the data content.

Standards for Adaptive and Bayesian Trial Designs

8.1.1 Specify Planned Adaptations and Primary Analysis

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

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

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

8.1.2 Evaluate Statistical Properties of Adaptive Design

While not necessary for simple designs, the statistical properties of complex adaptive clinical trial designs should be thoroughly investigated over the relevant range of important parameters or clinical scenarios (e.g., treatment effects, accrual rates, delays in the availability of outcome data, dropout rates, missing data, drift in participant characteristics over time, subgroup-treatment interactions, and/or violations of distributional assumptions). Statistical properties to be evaluated should include type I error, power, and sample size distributions, as well as the precision and bias in the estimation of treatment effects. Additional performance metrics may also be evaluated (e.g., the frequency with which specific adaptations occur, the likelihood of substantial covariate imbalance, the likely adequacy of final data for subgroup and safety analyses).

The programming code used to create the simulations should be retained with version control. The programming code and software used should be made available to stakeholders who have a need to know, including reviewing agencies.

8.1.3 Specify Structure and Analysis Plan for Bayesian Adaptive Randomized Clinical Trial Designs

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

8.1.4 Ensure Clinical Trial Infrastructure Is Adequate to Support Planned Adaptation(s)

The clinical trial infrastructure, including centralized randomization, data collection related to the assessment and recording of key outcomes, data transmission procedures, and processes for implementing the adaptation (e.g., centralized, web-based randomization), must be able to support the planned trial. In simple adaptive trials, qualitative verification of the capabilities of the proposed trial infrastructure may be adequate. Trials with more complicated requirements such as frequent interim analyses require thorough testing prior to trial initiation. Such testing should involve the trial's data collection and data management procedures, the implementation of the adaptive algorithm, and methods for implementing the resulting adaptation(s). The impact on the trial's operating characteristics of delays in collecting and analyzing available outcome data should be assessed.

8.1.5 Use the CONSORT statement, with Modifications, to Report Adaptive Randomized Clinical Trials

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

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

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

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

8.2.1 Describe Data Linkage Plans, if Applicable

For studies involving linkage of registry data with another data source, describe the other data source and its appropriateness and limitations for addressing specific hypotheses. Consider any additional requirements that may influence successful linkage, such as information needed to match patients, selection of data elements, and definitions used.

8.2.2 Plan Follow-up Based on the Registry Objective(s)

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

8.2.3 Describe Data Safety and Security

Research proposals and protocols should describe data use agreements, informed consent, data security, and approaches to protecting security including risk of re-identification of patients. If using previously collected data, describe how these address the risk of re-identification of patients and the actual use of data compared with the originally designed and consented use of the data.

8.2.4 Take Appropriate Steps to Ensure Data Quality

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

8.2.5 Document and Explain Any Modifications to the Protocol

Modifications to a registry protocol may be necessary for a variety of reasons. When modifications are necessary, document and explain any modifications to the formal study protocol.

8.2.6 Collect Data Consistently

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

8.2.7 Enroll and Follow Patients Systematically

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

8.2.8 Monitor and Take Actions to Keep Loss to Follow-up to an Acceptable Minimum

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

8.2.9 Use Appropriate Statistical Techniques to Address Confounding

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

Standards for Studies of Diagnostic Tests

8.3.1 Specify Clinical Context and Key Elements of Diagnostic Test Study Design

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

8.3.2 Study Design Should be Informed by Investigations of the Clinical Context of Testing

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

8.3.3 Assess the Effect of Factors Known to Affect Diagnostic Performance and Outcomes

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

8.3.4 Structured Reporting of Diagnostic Comparative Effectiveness Study Results

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

8.3.5 Give Preference to Randomized Designs of Studies of Test Outcomes

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

Appendix B: Recommendations

Recommended Actions

Topic	Action
Standards for Formulating Research Questions	• The Methodology Committee recommends that PCORI develop policies to encourage public registration of all PCORI studies and the sharing of study protocols, statistical code, and data. • Form a standing committee within PCORI to recommend appropriate methods for data sharing and to ensure that proper scientific credit is given to those sharing protocols, code, and data. • To speed implementation of standards in funding announcements, peer review, and other internal processes, PCORI staff should develop or have developed templates for the preparation and review of proposals that incorporate the key elements of the standards. Because some standards apply only to certain types of studies, a portfolio of templates applicable to various study designs should be developed.
Patient-Centeredness	• Support training in patient engagement methods for investigators and patient informants. • Improve the patient-reported outcomes (PRO) evidence base by supporting research on methods for assessing measurement properties (based on qualitative and quantitative evaluations), score interpretability, meaningfulness of score changes, and strategies for minimizing and interpreting missing PRO data in PCOR. • Evaluate patient dissemination activities, and require incorporation in future research of relevant learnings from this evaluation.
Research Prioritization	• Work closely with the Methodology Committee to build on previous work and implement the framework efficiently. • Base all PCORI targeted funding opportunity announcements on evidence gap analysis. • Require that applicants demonstrate how their proposed research fills a research gap for non-targeted funding opportunity announcements. • Support education and training activities to broaden the base of individuals prepared to apply and evaluate VOI. • Maintain peer review processes that avoid interference of participants and stakeholders with potential conflicts of interest. Peer review should incorporate patient perspectives.

General and Crosscutting Methods for All PCOR	• Sponsor randomized trials alongside registry studies to compare the validity of different methods for reducing confounding and bias. • Develop and disseminate software needed for sensitivity analyses and approaches to evaluating the assumptions underlying complex analyses such as instrumental variable analyses. • Develop and distribute software to reduce barriers that inhibit the use of more rigorous methods for handling missing data. • Provide training in methods for systematic reviews, modeling, and addressing missing data. • Promote accumulation of evidence to supplement common practices as guidance for future development of data network structures. • Promote approaches to privacy protection in data networks that also consider how to enhance data utility. • Require that missing and loss-to-follow-up data be reported in all research results.
Design-specific Methods	• Develop additional guidance specific to adaptive trials. • Support development and use of software for adaptive trials that can simulate complex designs. • Broaden experience with adaptive trials for PCOR, perhaps through funding of a cohort of adaptive trials on priority topic areas. • Mentor investigators and develop a "how-to" guide and a forum to share experiences with adaptive trials. • Develop coursework and other training opportunities for statisticians and other methodologists interested in developing expertise in adaptive trials. • Sponsor an Institute of Medicine committee to develop standards for research on medical tests.

Research Recommendations

Topic	Recommended Research
Patient-Centeredness	• Support research to develop a standardized nomenclature for patient engagement methods. • Create an infrastructure to support research on patient engagement. To facilitate this, PCORI should: • Develop a standardized nomenclature for patient engagement methods. • Develop a sample patient engagement plan to demonstrate the key elements required for patient engagement in the research process. The sample plan should illustrate engagement of both patient informants and study participants. • Systematically collect information about patient engagement methods from PCORI-sponsored studies. • Evaluate the effectiveness of patient informant engagement. • Synthesize results across studies. • Disseminate findings to improve patient engagement in PCOR
Research Prioritization	• Encourage intra- and extramural research in the development and practical application of VOI methods for PCOR, including through studies that examine the contribution of VOI methods to research prioritization when used in conjunction with other approaches to research prioritization. • Support empirical research to assess and improve research prioritization methods for use by PCORI. • Support extra- and/or intramural research to establish a best practice approach to consultative and collaborative patient engagement in topic generation that is suitable for the heterogeneity of the U.S. patient population. • Study the employment of research gap analysis to continue to develop the empirical evidence on its use. • Encourage studies, ideally with experimental designs, that assess different methods for engaging patients with diverse views and preferences and funneling their input into the peer review process in a consultative manner.

General and Crosscutting Methods

- Fund research on innovative ways to identify and recruit new users of treatment for research studies.
- Fund research on ways to identify and include reasonable treatment alternative comparators.
- Develop and disseminate templates for describing who is in each analysis and the potential sources of selection bias.
- Develop and disseminate methods for adequate analysis of data in cases where the treatment/exposure varies over time and it is not possible to adhere to these standards.
- Incorporate evolving new technology, such as the use of cloud technology, into ongoing work in the design of networks.
- Fund research on the best way to harmonize data elements across sources.
- Develop methods guidance for analyses for HTE in observational studies.
- Develop methods guidance on the use of Bayesian methods in HTE analyses and appropriate outcome scale for HTE analysis (e.g., risk difference, risk ratio, log of odds-ratio).
- Support the development of both analytic approaches and guidance for predictive approaches to HTE as well as for SGA with a focus on their use for PCOR.
- Develop methods guidance for HTE analyses in comparative effectiveness trials; the literature on HTE almost exclusively discusses use in placebo-controlled trials.

Design-specific Methods

For registry studies:

- Develop analytic techniques for addressing measured and unmeasured confounding.
- Develop analytic techniques for handling missing data that can be used in registry studies.
- Develop improved strategies for linking data while maintaining privacy protections and assuring that link data do not lead to re-identification in de-identified data.
- Develop innovative ways to reduce loss to follow-up as registries encompass longer time periods.

Appendix C: Patient Engagement

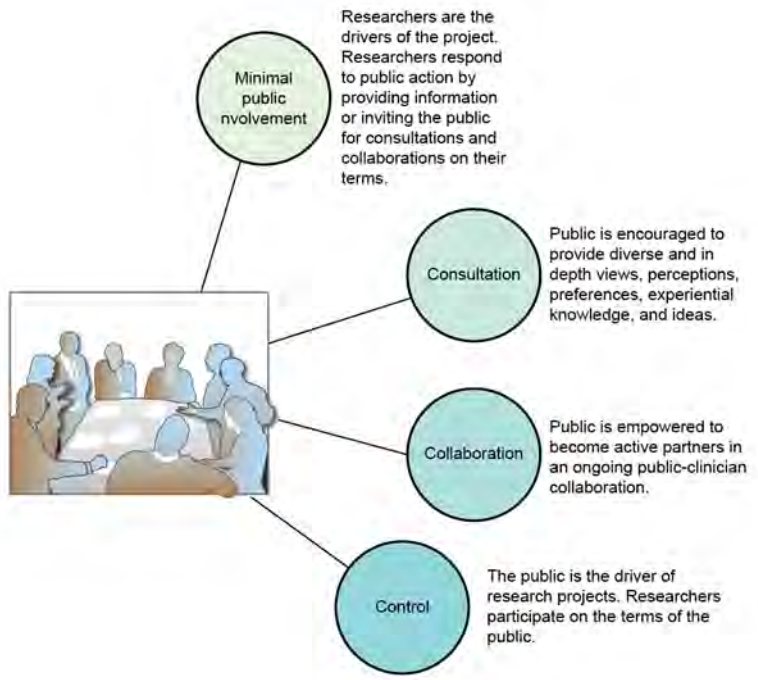
This appendix presents some of the Committee's current thinking related to patient engagement and patient centered outcomes. The review is not exhaustive, but presents some of the key concepts that the Committee expects to address in the coming months.

Patient Engagement in Health Research

Patients often want to know which treatments are best for their condition, given their personal characteristics. Clinicians share this concern. To make sure that "…selected topics truly represent a potentially large impact on the clinical or other outcomes that matter most to patients," (1) the Agency for Healthcare Research and Quality (AHRQ) and the Institute of Medicine recognize the need for meaningful involvement of patients as stakeholders in the selection and refining of topics for CERs. (2-4)

Patient engagement can be seen as a continuum, from passive patient involvement as study participants to active engagement as collaborators or research leaders. (5-8) A variety of models of engagement have been described. Arnstein (1969) illustrated a hierarchy of involvement with treatment models encompassing different forms of power sharing and partnership. (9) This ranges from citizen control at one end, forms of tokenism—including placation—in the intermediate position , and nonparticipation and, in worst cases, manipulation at the other end. Rowe and Frewer (2005) present a different model of engagement based on direction of communication: there can be one-way communication from research sponsors to the public; there can be one-way communication from public representatives to research sponsors; and there can be bi-directional communication to optimize engagement. (10) Oliver and colleagues (2008) categorize public/researcher interaction based on who initiates the engagement, the number of people who are engaged, and the social form of engagement. (6) (See Nass et al. 2012.) Figure C.1 displays a typology for patient engagement exemplified in these models.

Figure C.1 Typology of Patient Engagement in Health Research



Principles for Engaging Patients in Healthcare Research

In a background paper developed to support this report, the contractor's report (*Methods for Involving Patients in Topic Generation* http://www.pcori.org/what-we-do/methodology/) lists a number of principles for engaging the public in health care research that are appropriate regardless of the format used for patient engagement. These principles include:

Members of the public are empowered to become active, respected participants in the project.

- The engagement process is transparent and includes a conflict-of-interest statement.
- The process used to invite and select participants is inclusive and balanced in terms of ethnicity, gender, age, disease burden, and socioeconomic status.
- The roles and relationships for researchers and lay participants are clarified at the beginning of each project.

- The public is engaged using appropriate, validated, and diverse methods by staff experienced in PCOR or patient-centered care.

- The process is sustainable and establishes a culture of improvement. There are measures for quality control of patient participation to ensure that the integrity of the process of patient involvement is maintained over time and across different projects.

Methods for Engaging Patients in Health Research

The steps in the engagement process are choosing individuals to participate ("informant selection"), building reciprocal relationships, co-learning, and reassessment/feedback (*Eliciting Patient Perspective in Patient-Centered Outcomes Research – A Meta Narrative Systematic Review* http://www.pcori.org/what-we-do/methodology/). "Building relationships" may involve eliciting stakeholder perspectives and preparing stakeholders for engagement; "co-learning" involves communication and aspects of relationship building; and "reassessment and feedback" includes dissemination activities, evaluation of the engagement process, and ongoing communication between the relevant communities and researchers (*Eliciting Patient Perspective in Patient-Centered Outcomes Research – Expert Interviews Parts 1-4* http://www.pcori.org/what-we-do/methodology/).

There are many case studies of the value of patient involvement in designing, conducting, and disseminating research. Some of these describe experiences involving patients and caregivers in the design of specific research studies, often clinical trials. (11) Others use a community-based participatory research approach (CBPR). Many of the public-clinician partnerships formed based on a CBPR model were created to engage the public in health care services research with the goal of decreasing health disparities. (12, 13)

In these case studies, involving informants can both help shape research questions so they more closely match the questions patients ask (14) and build trust, which, in turn, may improve recruitment rates and retention of study participants. (6, 11, 15-22) (Nass P et al. 2012) Patient and community engagement can also strengthen the content and construct validity of measures. (23, 24)

Although models for engagement in PCOR are scant, related experiences with community-based participatory research (CBPR) are instructive.

Use of Patient-Reported Outcome Measures in PCOR

PCOR requires the use of study outcomes that matter to patients, outcomes that they notice and find meaningful. Direct reports by patients, captured through patient-reported outcome (PRO) measures of symptoms, function, and other effects of health conditions help ensure that research addresses questions that are important to patients and that inform decision-making.

The use of PRO measures pre-dates the current focus on PCOR and derives from multiple disciplines, including the field of educational assessment, where many advances in "psychometrics" –measuring the unobservable human experience–originated (see Jones and Thissen, 2007 for a review). (25)

PRO measure development has advanced over the last 20 years, with growing recognition of the value of direct patient input, and as modern test theory methods such as adaptive testing have been brought to the health assessment field.

Development of PROs requires patient input. Establishing content validity by ensuring that a proposed measure captures what patients actually think and feel is a critical early step in the development process. Psychometric properties to evaluate when determining whether a specific PRO is appropriate for a given study include reliability (internal consistency and test-retest), (26-28) validity (content, construct, criterion-related), (29, 30) responsiveness (sensitivity to change), (31-39) interpretability of scores and change in scores (i.e., clinically meaningful differences), (32, 34, 36-40) respondent and administrative burden, (41, 42) comparability of different assessment modes (paper, computer, interviewer-administered), (43-45) and cultural and language translations. (46-53)

Several published guidelines for development, evaluation, and use of PROs exists, most dating from the last decade; (28, 29, 31, 32, 43, 44, 54-66) guidelines for PRO use in CER are currently under development (Reeve, personal communication 17April 2012)

While several guidance documents indicate threshold values for acceptable levels of various psychometric properties, the Methodology Committee recognizes evidence gaps for many existing measures and also recognizes the time and resource requirements associated with collecting psychometric data for new measures. Researchers should acknowledge limitations in existing psychometric data but PRO use should not be precluded due to lack of psychometric data.

Appendix C References

1. Whitlock E, Lopez S, Chang S. Identifying, selecting, and refining topics. In: Agency for Healthcare Research and Quality. Methods Guide for Comparative Effectiveness Reviews. Rockville, MD2009.

2. Institute of Medicine (IOM). Consensus Report. Knowing What Works in Health Care: A Roadmap for the Nation. 2008 [cited 2012 January 10]; Available from: http://www.iom.edu/Reports/2008/Knowing-What-Works-in-Health-Care-A-Roadmap-for-the-Nation.aspx.

3. Tunis SR, Benner J, McClellan M. Comparative effectiveness research: Policy context, methods development and research infrastructure. Statistics in medicine. 2010 Aug 30;29(19):1963-76.

4. Whitlock EP, Lopez SA, Chang S, Helfand M, Eder M, Floyd N. AHRQ series paper 3: identifying, selecting, and refining topics for comparative effectiveness systematic reviews: AHRQ and the effective health-care program. J Clin Epidemiol. 2010 May;63(5):491-501.

5. Shippee NJ, Montori V. Patient and Public Involvement (PPI) in the Research Process. Rochester: Mayo Clinic; 2011:19.

6. Oliver SR, Rees RW, Clarke-Jones L, Milne R, Oakley AR, Gabbay J, et al. A multidimensional conceptual framework for analysing public involvement in health services research. Health Expect. 2008 Mar;11(1):72-84.

7. Happell B, Roper C. Consumer participation in mental health research: articulating a model to guide practice. Australas Psychiatry. 2007 Jun;15(3):237-41.

8. National Institutes of Health (NIH). Principles of Community Engagement (2nd ed). NIH Publication No. 11-7782. Bethesda, MD: NIH; 2011.

9. Arnstein S. A ladder of citizen participation. J Am Plann Assoc. 1969;35(4):216-24.

10. Rowe G, Frewer L. Science Technology Human Values. Spring 2005;30:251-90.

11. Edwards V, Wyatt K, Logan S, Britten N. Consulting parents about the design of a randomized controlled trial of osteopathy for children with cerebral palsy. Health Expect. 2011 Dec;14(4):429-38.

12. Cornwall A, Jewkes R. What is participatory research? Soc Sci Med. 1995 Dec;41(12):1667-76.

13. Viswanathan M, Ammerman A, Eng E, Gartlehner G, Lohr K, Griffith D, et al. Community-Based Participatory Research: Assessing the Evidence. Evidence Report/Technology Assessment No. 99 (Prepared by RTI-University of North Carolina Evidence-based Practice Center under Contract No. 290-02-0016). AHRQ Publication 04-E022-2. Rockville, MD: Agency for Healthcare Research and Quality. July 2004.

14. Chalmers I. What do I want from health research and researchers when I am a patient? BMJ. 1995 May 20;310(6990):1315-8.

15. Decker M, Hemmerling A, Lankoande F. Women front and center: the opportunities of involving women in participatory health research worldwide. J Womens Health (Larchmt). 2010 Nov;19(11):2109-14.

16. Entwistle VA, Renfrew MJ, Yearley S, Forrester J, Lamont T. Lay perspectives: advantages for health research. BMJ. 1998 Feb 7;316(7129):463-6.

17. Faulkner ARJ, 2012., from Changing Our Worlds: examples of user-controlled research in action. 2010, October [cited 2012 January 10]; Available from: http://conres.co.uk/pdfs/6823_INVOLVE_UCRCS_brochure_WEB.pdf.

18. Boote J, Baird W, Sutton A. Public involvement in the systematic review process in health and social care: a narrative review of case examples. Health Policy. 2011 Oct;102(2-3):105-16.

19. Patient Partner. The active involvement of Patients and Patient organisations in (the) clinical trial (development process). 2012; Available from: http://www.patientpartner-europe.eu/en/resources/active-involvement-in-the-process.

20. Chenoweth L, Kilstoff K. Facilitating positive changes in community dementia management through participatory action research. Int J Nurs Pract. 1998 Sep;4(3):175-88.

21. Rhodes P, Nocon A, Wright J, Harrison S. Involving patients in research: setting up a service users' advisory group. Journal of management in medicine. 2001;15(2):167-71.

22. Staniszewska S, Jones N, Newburn M, et al. User involvement in the development of a research bid: barriers, enablers and impacts. Health Expect. 2007;10:173-83.

23. Cashman SB, Adeky S, Allen AJ, 3rd, Corburn J, Israel BA, Montano J, et al. The power and the promise: working with communities to analyze data, interpret findings, and get to outcomes. Am J Public Health. 2008 Aug;98(8):1407-17.

24. Cotterell P. Exploring the value of service user involvement in data analysis: 'Our interpretation is about what lies below the surface'. Educational Action Research. Mar 2008;16(1):5-17.

25. Jones L, Thissen D. A history and overview of psychometrics. In: Rao C, Sinharay S, editors. Handbook of Statistics. The Netherlands: Elsevier; 2007. p. 1–27.

26. Kottner J, Audige L, Brorson S, Donner A, Gajewski BJ, Hrobjartsson A, et al. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. J Clin Epidemiol. 2011 Jan;64(1):96-106.

27. Frost MH, Reeve BB, Liepa AM, Stauffer JW, Hays RD. What is sufficient evidence for the reliability and validity of patient-reported outcome measures? Value Health. 2007 Nov-Dec;10 Suppl 2:S94-S105.

28. Scientific Advisory Committee of the Medical Outcomes Trust. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. [Research Support, Non-U.S. Gov't Research Support, U.S. Gov't, P.H.S. Review]. May 2002;11(3):193-205.

29. Magasi S, Ryan G, Revicki D, Lenderking W, Hays RD, Brod M, et al. Content validity of patient-reported outcome measures: perspectives from a PROMIS meeting. Qual Life Res. Epub ahead of print: Aug 26, 2011.

30. Streiner D. A checklist for evaluating the usefulness of rating scales. The Canadian Journal of Psychiatry / La Revue canadienne de psychiatrie. 1993;38(2):140-8.

31. Angst F. The new COSMIN guidelines confront traditional concepts of responsiveness. BMC Med Res Methodol. 2011;11:152; author reply

32. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008 Feb;61(2):102-9.

33. Revicki DA, Cella D, Hays RD, Sloan JA, Lenderking WR, Aaronson NK. Responsiveness and minimal important differences for patient reported outcomes. Health and quality of life outcomes. 2006;4:70.

34. Hays RD, Woolley JM. The concept of clinically meaningful difference in health-related quality-of-life research. How meaningful is it? Pharmacoeconomics. 2000 Nov;18(5):419-23.

35. Hays RD, Hadorn D. Responsiveness to change: an aspect of validity, not a separate dimension. Qual Life Res. 1992 Feb;1(1):73-5.

36. Wyrwich K, Norquist J, Lenderking W, Acaster S. Industry Advisory Committee of International Society for Quality of Life Research (ISOQOL). Methods for interpreting change over time in patient-reported outcome measures. Not published yet. 2012.

37. Kemmler G, Zabernigg A, Gattringer K, Rumpold G, Giesinger J, Sperner-Unterweger B, et al. A new approach to combining clinical relevance and statistical significance for evaluation of quality of life changes in the individual patient. J Clin Epidemiol. 2010 Feb;63(2):171-9.

38. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol. 2003 May;56(5):395-407.

39. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991 Feb;59(1):12-9.

40. Sprangers MA, Moinpour CM, Moynihan TJ, Patrick DL, Revicki DA, Clinical Significance Consensus Meeting G. Assessing meaningful change in quality of life over time: a users' guide for clinicians. Mayo Clinic proceedings Mayo Clinic. 2002 Jun;77(6):561-71.

41. Turner RR, Quittner AL, Parasuraman BM, Kallich JD, Cleeland CS, Mayo FDAP-ROCMG. Patient-reported outcomes: instrument development and selection issues. Value Health. 2007 Nov-Dec;10 Suppl 2:S86-93.

42. Scientific Working Group Quality of Life and Symptoms. Scientific Working Group Quality of Life and Symptoms. European Hematology Association; Available from: http://www.gemclinic.ru/konf01eng.php.

43. Snyder CF, Aaronson NK, Choucair AK, Elliott TE, Greenhalgh J, Halyard MY, et al. Implementing patient-reported outcomes assessment in clinical practice: a review of the options and considerations. Qual Life Res. 2011 Nov 3.

44. Coons SJ, Gwaltney CJ, Hays RD, Lundy JJ, Sloan JA, Revicki DA, et al. Recommendations on evidence needed to support measurement equivalence between electronic and paper-based patient-reported outcome (PRO) measures: ISPOR ePRO Good Research Practices Task Force report. Value Health. 2009 Jun;12(4):419-29.

45. Kongsved SM, Basnov M, Holm-Christensen K, Hjollund NH. Response rate and completeness of questionnaires: a randomized study of Internet versus paper-and-pencil versions. J Med Internet Res. 2007;9(3):e25.

46. Gawlicki M, McKown S, Handa M. Preempting difficulties in liniguistic validation: the use of face validation to create more sound translations. International Society for Pharmacoeconomics Outcomes Research (ISPOR); Nov 6-9, 2011; Prague, Czech Republic.

47. Hagell P, Hedin PJ, Meads DM, Nyberg L, McKenna SP. Effects of method of translation of patient-reported health outcome questionnaires: a randomized study of the translation of the Rheumatoid Arthritis Quality of Life (RAQoL) Instrument for Sweden. Value Health. 2010 Jun-Jul;13(4):424-30.

48. Acquadro C, Conway K, Hareendran A, Aaronson N. Literature review of methods to translate health-related quality of life questionnaires for use in multinational clinical trials. Value Health. 2008 May-Jun;11(3):509-21.

49. Maneesriwongul W, Dixon JK. Instrument translation process: a methods review. J Adv Nurs. 2004 Oct;48(2):175-86.

50. Sperber AD. Translation and validation of study instruments for cross-cultural research. Gastroenterology. 2004 Jan;126(1 Suppl 1):S124-8.

51. Schmidt S, Bullinger M. Current issues in cross-cultural quality of life instrument development. Arch Phys Med Rehabil. 2003 Apr;84(4 Suppl 2):S29-34.

52. Hutchinson A, Bentzen N, Konig-Zahn C. Cross cultural health outcome assessment - a user's guide. Ruinen, The Netherlands: European Research Group on Health Outcomes; 1997.

53. Ware JE, Jr., Keller SD, Gandek B, Brazier JE, Sullivan M. Evaluating translations of health status questionnaires. Methods from the IQOLA project. International Quality of Life Assessment. Int J Technol Assess Health Care. 1995 Summer;11(3):525-51.

54. Mokkink LB, Terwee CB, Gibbons E, Stratford PW, Alonso J, Patrick DL, et al. Inter-rater agreement and reliability of the COSMIN (COnsensus-based Standards for the selection of health status Measurement Instruments) checklist. BMC Med Res Methodol. 2010;10:82.

55. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. Protocol of the COSMIN study: COnsensus-based Standards for the selection of health Measurement INstruments. BMC Med Res Methodol. 2006;6:2.

56. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: a clarification of its content. BMC Med Res Methodol. 2010;10:22.

57. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. Qual Life Res. 2010 May;19(4):539-49.

58. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010 Jul;63(7):737-45.

59. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. Qual Life Res. 2012 May;21(4):651-7.

60. Johnson C, Aaronson N, Blazeby J, Bottomley A, Fayers P, Koller M, et al. EORTC Quality of Life Group: Guidelines for Developing Questionnaire Modules. 4th. 2011 [cited 2011 November 26]; Available from: http://groups.eortc.be/qol/Pdf%20presentations/Guidelines%20for%20Developing%20questionnaire-%20FINAL.pdf.

61. Cella D. Manual of the Functional Assessment of Chronic Illness Therapy (FACIT) measurement system. Evanston, IL: Northwestern University; 1997.

62. Rothman M, Burke L, Erickson P, Leidy N, Patrick D, Petrie C. Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR Good Research Practices for Evaluating and Documenting Content Validity for the Use of Existing Instruments and Their Modification PRO Task Force Report. Value Health. 2009 Nov-Dec;12(8):1075-83.

63. Wild D, Eremenco S, Mear I, Martin M, Houchin C, Gawlicki M, et al. Multinational trials-recommendations on the translations required, approaches to using the same language in different countries, and the approaches to support pooling the data: the ISPOR Patient-Reported Outcomes Translation and Linguistic Validation Good Research Practices Task Force report. Value Health. 2009 Jun;12(4):430-40.

64. Valderas JM, Ferrer M, Mendivil J, Garin O, Rajmil L, Herdman M, et al. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. Value Health. 2008 Jul-Aug;11(4):700-8.

65. Dewolf L, Koller M, Velikova G, Johnson C, Scott N, Bottomley A. EORTC Quality of Life Group: Translation Procedure. 3rd. 2009 [cited 2011 November 26]; Available from: http://groups.eortc.be/qol/downloads/translation_manual_2009.pdf.

66. Wild D, Grove A, Martin M, et al. Principles of good practice for the translation and cultural adaptation process for patient-reported outcomes (PRO) measures: report of the ISPOR Task Force for Translation and Cultural Adaptation. Value Health. 2005;8:94-104.

Appendix D: Translation Table

Appendix D-1
REQUEST FOR INFORMATION—Input on a Draft Translation Table Framework (*This RFI was open January 3-February 17, 2012*)

BACKGROUND AND PURPOSE

The Patient-Centered Outcomes Research Institute (PCORI) is an independent, non-profit research organization created by the Patient Protection and Affordable Care Act of 2010 (PPACA). The mission of PCORI is to help people make informed health care decisions – and improve health care delivery and outcomes – by producing and promoting high-integrity, evidence-based information derived from research guided by patients, caregivers and the broader health care community. Research commissioned by PCORI aims to be responsive to the values and interests of patients and provide patients and those who care for them with reliable, evidence-based information for the health care choices they face.

The Methodology Committee of PCORI was established to develop and improve the science and methods of comparative clinical effectiveness research. In particular, the legislation calls on the Methodology Committee to develop "a translation table that is designed to provide guidance and act as a reference for the Board to determine research methods that are most likely to address each specific comparative clinical effectiveness research question," and to produce a report by May 2012 outlining the progress on the development of this translation table.

The Methodology Committee has developed a preliminary translation framework that will inform the development of the translation table. The purpose of this Request for Information (RFI) is to invite input on the translation framework components and to engage stakeholder communities in the development of this translation tool. Responses to this RFI will be considered for inclusion in the May 2012 report. Details on the information requested and response submission instructions are provided in later sections of this document.

INTRODUCTION TO THE PROPOSED TRANSLATION FRAMEWORK AND TOOL

Box 1 defines patient-centered outcomes research, translation table, translation framework, and translation tool within the context of this RFI. **Figure 1** diagrams the proposed structure of the translation tool**. Box 2** describes the background and elements of a patient-centered research question. **Box 3** describes the proposed translation framework components.

Box 1: Definition of Patient-Centered Outcomes Research, Translation Table, Translation Framework, and Translation Tool

Patient-Centered Outcomes Research (PCOR): PCOR aims to help people make informed health care decisions and allows their voice to be heard in assessing the value of health care options. PCOR:

- Assesses the benefits and harms of preventive, diagnostic, therapeutic, or health delivery system interventions to inform decision-making, highlighting comparisons and outcomes that matter to people;
- Is inclusive of an individual's preferences, autonomy and needs, focusing on outcomes that people notice and care about such as survival, function, symptoms, and health-related quality of life;
- Incorporates a wide variety of settings and diversity of participants to address individual differences and barriers to implementation and dissemination; and
- Investigates (or may investigate) optimizing outcomes while addressing burden to individuals, resources, and other stakeholder perspectives.

Translation Table (according to PPACA): "The translation table will provide guidance and act as a reference for the PCORI Board to determine research methods that are most likely to address specific comparative clinical effectiveness research questions."

Translation Framework: The *translation framework* provides the theoretical underpinning and organizing structure for the t*ranslation table and translation tool* that will help users identify the range of appropriate research designs and analytic approaches to answer specific patient-centered research questions. It is defined by a set of *framework components* (see Box 3) that can be used to guide the user in making choices in study design and analytic methods based on current scientific knowledge.

Translation Tool: Although the legislative mandate is to create a translation table, the usefulness of a tabular format may be limited. A *translation tool* is a dynamic implementation of the *translation framework* to help users apply it to specific research questions. Such a tool might take the form of decision trees, weighting or optimization algorithms, or other formats. More than one design recommendation may result, along with designs that are deemed unacceptable.

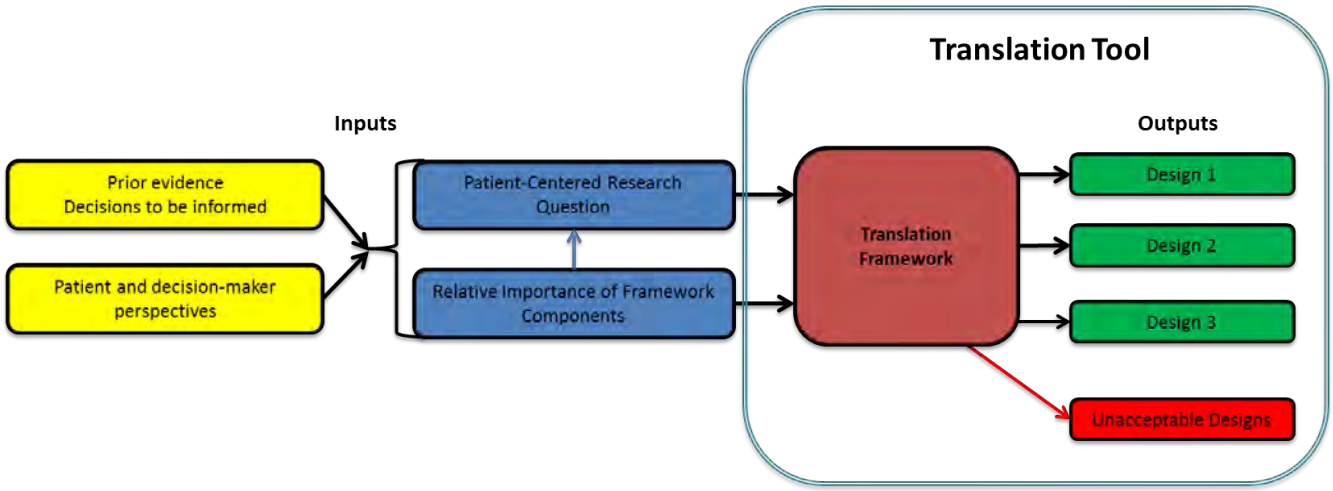
Proposed Structure of the Translation Tool

The translation tool will be the instrument by which the translation framework is operationalized. The patient perspective, patient-centered research question, and the relative importance of framework dimensions are key inputs to the translation tool.

The steps preceding the development of the research question establish how much and what kind of evidence is needed. These include a summary of prior studies, including what is known, unknown and why, and the decision the study in question is supposed to inform. Multiple perspectives can affect the choice of design; perspectives include those of patients, clinicians, researchers and policy makers. These perspectives affect choice of study design through both the choice of the patient-centered research question (see Box 2) and the relative importance patients and decisions-makers place on the various framework components (see Box 3).

Based on the question and relative importance of the components of the translation framework, the translation tool provides an output outlining acceptable and unacceptable designs. Output could be a ranked list of potential designs, a list of recommended and not recommended designs, or a table of potential designs with a description of the trade-offs involved with each design.

Figure 1: Proposed Structure of the Translation Tool.



Defining a Patient-Centered Research Question

A key step in designing patient-centered research is defining a research question that addresses questions that are important to patients and persons caring for them. PCOR should answer questions like:

- "Given my personal characteristics, conditions and preferences, what should I expect will happen to me?"
- "What are my options and what are the benefits and harms of those options?"
- "What can I do to improve the outcomes that are most important to me?"
- "How can the health care system improve my chances of achieving the outcomes I prefer?"

For example, patients seek research that involves populations with similar characteristics to themselves, compares the actual intervention and comparator they are considering, assesses outcomes important to them and their caregivers, and uses a time frame that captures relevant benefits and harms in a setting (e.g. hospital or clinic) typical of what they might find in their community.

Box 2. Defining a Patient-Centered Research Question

Background

- Prior evidence: Empirical studies of intervention-outcome and studies related to mechanism
- Decision(s) that the research is intended to inform

A patient-centered research question specifies the following elements:

- Population of patients / research participants
- Intervention(s) relevant to patients in target population
- Comparator(s) relevant to patients in target population
- Outcomes meaningful to patients in target population
- Timing: outcomes and length of follow-up
- Setting and providers

Proposed Translation Framework Components

Every design has characteristics that must be balanced when developing a study to address a particular research question. For example, in order for the results to be obtained faster or to maximize external validity, one might conduct a study with less internal validity than an RCT. A

study design without a comparator, based on information in a device registry, might be acceptable for assessing device failure rates, but not to assess device effectiveness. A question about the effectiveness of a moderately toxic cancer therapy might put highest priority on minimizing bias and maximizing precision, which may require an RCT. Often, logistical issues can be more important than scientific ones, e.g. if only a limited number of patients are available to study, or if the data in existing datasets is not well suited for the question. The elements that should be assessed to inform choices and tradeoffs that will be made in selecting research designs and methodologies are captured by these proposed *translation framework components*.

Box 3. Proposed Translation Framework Components

Intrinsic translation framework components

- Internal validity (bias)
- External validity (generalizability, or applicability to non-study settings and populations)
- Precision
- Heterogeneity in risk or benefit (e.g. subgroup or "personalized" evidence)
- Ethical dimensions of the study (including considerations of risk-benefit balance and study burden for study participants)

Extrinsic translation framework components

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

Prioritization of framework components is a key input into the translation framework and may affect the question itself. Policy makers, clinicians, researchers and patients may prioritize framework components differently. Hence, input is requested on how differing perspectives might be incorporated into the choice of study design.

The proposed framework components listed above may not apply perfectly to the wide range of research methods and questions comprising patient-centered outcomes research. The application of the translation framework to particular research domains may require customization or modification of the framework.

INFORMATION REQUESTED

We seek suggestions for the improvement of the draft framework components, thoughts on alternate components or frameworks, discussion of how these components apply to different research domains, and comments on the limitations of the proposed framework.

To fully explain and illustrate input and comments submitted, we encourage respondents to provide specific examples or case studies based on patient-centered research questions. Case studies should briefly describe the research question of interest and demonstrate application of the translation framework. The decisions about the design and methods used in patient-centered outcomes research require consideration of the tradeoffs in applying each translation framework component. Authors of case studies should discuss these decisions/tradeoffs from the perspective of at least one potential decision-maker or stakeholder.

Below is a partial list of broad research domains from which such examples or case studies might be drawn.

- Studies on the safety and/or effectiveness of medications
- Studies on the safety and/or effectiveness of vaccines
- Studies on the safety and/or effectiveness of devices
- Studies on the safety and/or effectiveness of behavioral therapies
- Studies comparing two different treatment modalities, e.g. medication vs. device, medication vs. behavioral therapy, surgery vs. physical therapy etc.
- Studies on the safety and/or effectiveness of surgical interventions
- Studies on the safety and/or effectiveness of diagnostic tests
- Studies on the safety and/or effectiveness of imaging strategies
- Studies on the effectiveness of health promotion programs
- Studies on the effectiveness of delivery system interventions
- Modeling the effectiveness of therapies for populations beyond primary evidence

We are not seeking case studies on how to develop patient-centered research questions but rather, discussion of the tradeoffs between designs one would make based on preferences on the relative importance of framework components. For example, for a research question involving use of a new medical device, one would discuss the different designs that would be most appropriate/ inappropriate based on prior evidence in this area and the relative importance of framework components (e.g., time urgency, precision, subgroup effects) to the decision-maker.

Case studies based on patient-centered research questions and supported by references and other

documentation have the greatest chance of being used in the report.

Responses to this RFI will be considered for inclusion in the May 2012 report. Authors of case studies chosen for inclusion in the published report or of suggestions that result in qualitative modification of the framework will be publicly acknowledged.

- HOW TO SUBMIT A RESPONSE
- Responses to this RFI must be submitted electronically [through http://www.pcori.org/] no later than **February 17, 2012 at 5:00 PM ET**.
- All responses must include the names of its author(s), organization affiliation(s), contact email address(es), phone number(s) and conflict of interest declaration to be considered for inclusion in the May 2012 Methodology Report. Authors may be contacted for additional information.
- INQUIRIES
- Questions about this RFI should be submitted electronically to RMWG@pcori.org.

ABOUT PCORI POLICIES

PCORI intends to acknowledge consenting authors of detailed examples and case studies that are incorporated, in whole or in part, into the Methodology Report. PCORI will contact authors selected for acknowledgment in advance of publication.

Conflict of Interest – PCORI requires disclosure of any potential conflicts of interest, to be considered for acknowledgement in the published report.

DISCLAIMER

Response to this RFI is voluntary. Authors are responsible for obtaining any necessary permission to submit a response to this request. PCORI does not intend to make any awards for funding based on responses to this RFI or to otherwise pay for the preparation of any information submitted or for PCORI's use of such information. PCORI reserves the right to use information provided by respondents for any purpose deemed necessary and legally appropriate. Any individual(s) or organization(s) responding to this request should ensure that its response is complete and sufficiently detailed.

RESPONDERS TO THE RFI

Isaac Ampomah; Chris Barker, PhD; Lorry Blath; Diana Brixner, PhD, RPh; Winifred Carson-Smith; Francesco Chiappelli; Dennis J. Cotter; Fred Cox, PhD; Kathleen "Casey" Croy; William Doucette, PhD; Andrea Douglas; Robert W. Dubois, MD, PhD; Richard Gliklich, MD; Gordon H. Guyatt, MD; Kelly Haenlein; Rajendra Jagad; Simon Kim, MD; William Lang; Vincent Lau, PhD; Sumene Li; Joanne Lynn; Daniel Malone, PhD; Newell McElwee; Penny Mohr, MA; Victor M. Montori, MD, MSc; Eileen Morrissey; Vinit Naire; Matthew Rousculp; Ryan Saadi, MD, MPH; Gloria E. Sarto MD, PhD; Nilay D. Shah, PhD; Sean Sullivan, PhD; Jeffery Talbert, PhD; Mae Thamer; Carolyn Thorpe, PhD; Kerry Willis; Richard Willke; Yi Zhang.

Appendix D-2
SUMMARY OF RFI RESPONSES

Response	Focus	Comments
1	Studies of medical devices	RCTs may not be feasible for medical devices: • Devices frequently undergo product modifications over time. • Difficult to obtain patient consent for trials resulting in small sample sizes and shorter-term studies. • Difficulty in blinding. • Efficacy depends not only on the device but how it is used. • Competition limits comparative effectiveness studies of relevant devices. • Consider other study designs for devices including observational data, expert opinion, modeling techniques, and synthesis of studies. Case study: Drug eluting stents.
2	Case study	**Includes a revised translation tool schematic that diagrams the patient/stakeholder engagement process.** **Case study: Patient centered preoperative empowerment education.**
3	Translation table	• Add short- and long-term benefit and harm as outcomes (RFI page 2, Box 1, PCOR definition and sub bullet 4). • Consider the following edits (RFI page 4): • "Given my personal characteristics, conditions and preferences, what *health outcomes* should I expect, *both short-term and long-term*?" • "What are my options and what are the *likelihood* benefits and harms of those options, *under circumstances that are relevant to me*?" • "What can I do to improve the outcomes that are most important to me?" • "*What* can the health care system *do to* improve my chances of achieving the outcomes I prefer?" (Note: this revision aligns with the wording used in bullet 3) • The statement that, "[a] study design without a comparator, based on information in a device registry, might be acceptable for assessing device failure rates, but not to assess device effectiveness" (RFI pages 4 and 5) might be construed as setting a standard regarding registry data and utilization for exploring effectiveness. Registries can be a helpful source of effectiveness outcome evidence, especially when there are accepted definitions of the treatment goal.

Response	Focus	Comments
4	Intended use and relevance of translation framework/table/tool	Regarding intended use: • It is unclear how and by whom this tool is to be used. Is this for those applying for PCORI grant funding, will this be used to judge their grant submissions? • What is the purpose and value of the translation tool and by whom and how it would be used? • The framework definition implies that it will guide researchers in what methods to consider for patient centered research questions. • The framework would be used primarily by PCORI administrators when prioritizing funding opportunities. In addition, the framework could be used by reviewers and by grant applicants. Regarding relevance: • The framework is limited by a narrow perspective in use and reimbursement of health care resources; cost is a stark omission. • Under the proposed structure of the tool, shouldn't developing the research question be the first step? • What is the implication if a researcher chooses a method that the tool deems unacceptable? • One approach would be that PCORI requires a section in grant applications that explicitly addresses patient and perhaps other perspectives on the framework components. • A second approach could be that researchers engage in work to gather perspectives of patients and decision makers regarding selected cases (study options for given research questions) across a range of patient decisions. • It would be beneficial to carefully define each component and possibly break out some of these categories. • It is unclear from the framework if a true Bayesian analysis is necessary or required prior to reaching the conclusion that future research is warranted. • The term "Translation Tool" seems misleading since this tool is really about approaches to study the question, not how to translate the findings to meaningful use by patients and providers. • This tool does not address how studies should be designed so that the results can be interpreted by patients and providers. • The current structure is too simplistic and doesn't address the attributes to be considered to determine acceptable and unacceptable study designs. • The tool is so generic that it gives maximum flexibility to PCORI to judge most things as being within scope of patient-centered outcomes research. Case studies: • Treatment choice decision tool for patients with CML. • How does a comprehensive medication review affect older adult medication use and outcomes? • What is the effect of continuing versus discontinuing medications for osteoporosis in bedridden patients with limited life expectancy on hospitalizations for fractures and health-related quality of life? • How will a proposed therapy/drug impact my disease (an oncology perspective)?

Response	Focus	Comments
5	Research on complex relationships	• Consider system improvement in complex situations, which is not merely the aggregation of tests but also the deliberate testing of hunches about causal chains. In these situations, RCTs are not the "gold standard," but are actually implausible or uninformative. Incorporate methods using realistic evaluation, control charts, tests of change, and welcoming of context elements where they are critically important. • PCORI projects should attend to costs, burdens, manpower, and roles, so that the likely effects of choices are known. • Framework does not account for context-dependent relationships between intervention and effects. Case study: Studies of care transitions.
6	Case study	Case study: Actualizing patient-centered outcomes research through a dental evidence-based decisions practice-based research network.
7	Subgroups in research	Regarding trade-offs and relative importance of framework components: In the decision on whether to use subgroups in RCTs versus develop a risk score in observational studies and then stratify RCT results by risk score one would weigh the trade-offs of internal validity (bias) and external validity (generalizability, or applicability to non-study settings and populations). Subgroups and interaction tests may provide greater internal validity within a study, but use of prediction tools may allow comparisons across studies thus facilitating assessment of applicability. The latter is probably more important in assessing outcomes in the Chronic Kidney Disease patient population.
8	Racial and ethnic minorities in research	To improve utilization of translational research: • Fund or develop mechanisms to fund baseline research of racial/ethnic communities by investigators of color. • Incorporate community into research methodologies addressing patient outcomes in minority communities. • Structure research to include ethnic/cultural criteria in research constructs.
9	Patient centric framework	Many patients, hospitals, and drug companies have designed their services and products based on a current model that is health care service provider centric. We need to help them slowly transition into a patient centric framework. • Need to introduce ambulatory health tracking using biotechnology and surveillance systems. • Increase the degree of freedom to our doctors so they can try other alternatives and figure out better ways to measure their successful interventions than drug prescriptions. • Measure environmental variables and drug interaction using physiological stress responses of the human body. • Design better ways to account for all variables and mapping events that influence latency in health outcomes.

Response	Focus	Comments
10	Translation table	• Finalize the working definition of "patient centered outcomes research" before proceeding to finalize the methodological standards and translation table and related tools. This should be done through an open and transparent process or, in order to avoid delaying the Committee's work, utilize the existing statutory definition of "comparative clinical effectiveness research" and related statutory language as the basis for the Methodology Committee's work. • Consider heterogeneity and patient preference as important elements of patient-centered research. • Continue to place emphasis on approaches for capturing "patient-centered" outcomes (e.g., patient reported outcomes and subpopulation differences) in different types of studies. • Replace the words "harms" and "safety" throughout the framework with the word "risk," which would ensure that research questions are consistent with PCORI's statutory requirement that calls for the Institute to conduct research assessing the "health outcomes and clinical effectiveness, risks and benefits" of two or more interventions. • Describe how the translation table and tool integrate with the research priorities and agenda being developed by PCORI. • The proposed structure of the translation tool should include input points for broad stakeholder engagement throughout the model and demonstrate a strong reliance on clinical expertise in order to identify the questions that have the most significant impact on patient care. • Develop research standards through an open, transparent process with adequate opportunities for public input.

Response	Focus	Comments
11	Translation table and framework	• In the proposed structure, outputs are listed as, "a ranked list of potential designs, a list of recommended and not recommended designs, or a table of potential designs with a description of the trade-offs involved with each design." The output may also include a "family" of studies that in combination may address specific stakeholder questions. A research "strategy" as opposed to a single research design may be required to produce answers that will be useful to a broad group of decision-makers. • Throughout the planning and pilot phase of research and as the study question is refined there should be a feedback loop back to the stakeholders to ensure the study continues to address the central concerns. • Add an intrinsic framework component assessing "Is real world (effectiveness question) vs. explanatory (efficacy question) evidence required?" • Explicitly include the "ethics of randomization" in the ethical dimensions of the study. • Include the "regret" or the cost associated with getting an incorrect answer in the extrinsic translation framework components. • Include the cost of research studies as an extrinsic factor that should be weighed in selecting appropriate study methods. • Consider including the following factors in the translation framework: • 1. Factors leading to a randomized design (relative efficacy, feasible and ethical to randomize, prognostic variables are unclear and most variation in outcomes is unexplained, biologic process of disease is not well understood, risk associated with an incorrect answer is high, modest anticipated differences in effect size). • 2. Factors leading to an observational design (relative effectiveness is of interest, effect size is relatively large and selection bias can be reasonably controlled, relatively rare adverse events, safety is most important outcome of interest, major focus is on adherence or compliance with therapy, interest in associations of outcomes by patient subgroups and observed practice patterns, risk associated with an incorrect answer is low, efficient use of resources or need is main interest). • Enhance methods for eliciting values from a broad set of stakeholders regarding their willingness to trade off some of the factors listed in Box 3, "Proposed Translation Framework Components," and obtaining consensus about priorities for the focus of a question. • Include subpopulations in the populations of patients/research participants. • Clarify the purpose of the translation tool and the degree to which flexibility is encouraged and allowed. • The optimal study methodology will depend upon the questions, evidence gaps, and outcomes of interest rather than the treatment comparisons. Case study: Delphi survey of research question and study design preferences.
12	Case studies	Case studies include: • Comparative effectiveness of surgery and radiation for localized prostate cancer focusing on modern technologies and controlling for differences in patients and treatments that may affect outcomes. • Compare the effectiveness of comprehensive care coordination programs, such as the medical home, and usual care in managing children and adults with severe chronic disease, especially in populations with known health disparities.

Response	Focus	Comments
13	Study designs	• The RFI does not clearly define a translation table or the framework. • While the idea of comparative evaluation research is relatively new, the study designs that may be expected to provide the relevant information or "evidence" are well known. These include double blind RCT, Bayesian adaptive clinical trials, and nonexperimental studies, including observational, registries, and meta-analyses of existing studies.
14	Translation table and framework	• Numerous factors affect the choices in study design and setting in complex ways that would be very difficult to distill into categorical recommendations. The diagrams and boxes in the document suggest a simplistic tool that would take the framework components and prioritization as inputs and create study design choices as output, but the decisions made when defining the study question and assessing design features are much more complicated than depicted. • A table may not be the appropriate output tool from this endeavor. • A well defined patient-centered research question is critical to define before investigating specific study designs, and this research question will depend in large part on the prior related evidence, decisions to be informed, and patient and decision-maker perspectives, among other factors. • The terminology doesn't clearly specify the types of decisions to be informed by CER evidence, the accountable parties for making the decision (decision makers), or the parties who have an interest in the decision but who are not the decision makers (stakeholders). This lack of clarity is continued in the framework itself. • How one prioritizes framework components is a matter of perspective, and different perspectives will likely weight them differently. What would be more useful would be to output several designs and give potentially a rating of how each of the framework components is addressed (high, medium, low) by each design. • There are a number of additional finer categories for the recommended framework components that would significantly impact the selection of design. • Good implementation of this tool would be to output multiple designs and give their pros and cons so that researchers can weigh those and make informed decisions. Ideally, the tool would allow a researcher to output all possible designs and see which are not at all feasible, and how each of the other designs address the components and limitations.

Response	Focus	Comments
15	Translation table and framework	• The purpose of the translation table is unclear and there is a lack of information on its elements. • It would be helpful to offer a more well-defined, thorough analysis of key principles of health services research as a foundation and provide additional information, such as the intended audience and its anticipated use(s). • Development of a complicated methodological framework when a key underlying element - the final definition of PCOR - has not yet been finalized and publicly released is ambitious. • There is little information on how researchers should consider and incorporate patient and decision-maker perspectives. • The framework does not seem to explicitly recognize that there may be a wide array of comparative effectiveness questions that a given researcher may seek to investigate. How will one framework be applicable to all research questions? • Needs elaboration of the research components in the framework (prior evidence, decision(s) the proposed work intends to inform, and perspectives of patients and decision-makers), and incorporate guidance on: how to identify each of the three components listed; how the three components relate to one another; and the relative weight each should receive in the PCOR process. • Needs explicit recommendations on how to evaluate and consider prior evidence when performing PCOR. • Add a key question concerning the reconciliation of conflicting evidence. PCORI can play a critical role in providing recommendations on how to weigh competing results in decision-making. • The concept of the decision to be informed can be broadened and clarified to suggest that researchers consider the intent of the research when designing the patient-centered research question. This be incorporated into all aspects of the translation framework, but clarify that it may not always be as specific as addressing a decision at the point of patient care. • In addition to elements aimed at specifying a research question (population, intervention, comparators, outcomes, timing, and settings), it may be useful for researchers to denote the expected value of the information in their research question. Include such individuals as caregivers, family members, and others who play a role in a patient's health care delivery and support. • While considerations of population, intervention, comparator, and outcomes (PICO) may align well with many study designs, it may not be the optimal framework to capture key components of less traditional approaches, such as N=1 case studies or pragmatic studies. • While the translation framework is fairly comprehensive in its overview of key intrinsic and extrinsic components of study designs, guidance on each of these components is needed: How each will be measured (e.g., the value of each), and the relative importance of each (e.g., how to weigh tradeoffs among these factors). • Incorporate the following components to enhance the clarity and comprehensiveness of the framework features: include reliability, or designing a study that is replicable, in the list of intrinsic translation framework components; provide greater clarity around the factors that should be considered related to precision in the intrinsic framework components; consider incorporating the anticipated cost of the study into the extrinsic translation framework components; include the value of the research in the extrinsic translation framework.

Response	Focus	Comments
15	Translation table and framework (continued)	• Incorporate the following domains for consideration when formulating studies: health-related quality of life; patient and caregiver satisfaction with treatment or the treatment process; sequencing of treatments; set of treatment options; diagnostic tests in conjunction with therapies that may require the testing; and use of models to predict health outcomes.
16	Using administrative claims data	• Healthcare administrative claims databases are one of the most reliable and cost effective resources for a variety of health outcomes research (epidemiology, drug safety, comparative effectiveness, therapeutic efficacy, disease burden, patient preferences and clinical outcomes). • A challenge in utilizing claims data is the relative variety and differences that exist even among the various Health Plan administrative claim datasets. • A reliable way of addressing these differences while leveraging the value of administrative claims data sets is by the use of a validated Common Data Model (CDM), a collection of data tables and data elements that has been conformed to a particular standard.
17	Translation table and framework	• While one of the goals of the document is to bring transparency to the process of determining study design, it is quite complex and difficult to understand, and may be difficult to apply to real PCOR questions. The terminology used in the document is confusing, often using very similar terms that share definitions that are difficult to distinguish from each other (e.g. “translation table,” “translation framework components,” etc.). • It appears that the point at which the translation tool will be applied, the patient centered research question has been determined. As such, the order is incorrect and the question should come first, followed by evaluation of prior evidence and patient/decision-maker perspectives. We recommend reversing the order of the blue and yellow boxes. • The translation tool is missing an explicit evaluation metric/rubric to assess the strengths and weaknesses and appropriateness of each study design based on the research question of interest. • Replication of results is important in evaluating research questions; this should be noted as a consideration for evaluating prior evidence and the need for additional studies in the inputs section. • In Box 2, defining a research question should only involve the bullets in the second half of the box; the first half is what needs to be considered to determine and fine-tune the research question for evaluation. • Box 3 is missing an existing evidence framework and should call out a set of criteria for evaluating the patient-centeredness of the study, e.g. deciding how a comparator is relevant to patients. • Consider omitting heterogeneity in risk or benefit (e.g. subgroup or “personalized” evidence), since this may not be an appropriate factor in evaluating types of study design.

Response	Focus	Comments
18	Translation table and framework	• Consider including "Interpretability" as an external framework component. • Consider adding "resource constraint" as an extrinsic translation framework. The value of new evidence to be generated needs to be considered relative to the opportunity cost of investing in other patient centered priorities and alternative study designs. • Internal validity also includes sensitivity analyses. Internal validity may be supported when different study designs, analytic methods, and sensitivity analyses within the same data set produce similar effect estimates. • Include other components of generalizability: reproducibility and consistency across studies. • Provide clarity regarding the scope of "precision." Narrowly defined, this component could be considered the measure of variability of an effect or predicted response. However, broader statistical considerations could be included (prespecified analysis plan, power, model building/assumptions/diagnostics, missing data, multiple comparisons, sensitivity analysis, etc.) • A tradeoff not explicitly included in the framework is the target of inference that is important to the stakeholder. A study designed to assess which treatment is best for the individual patient may not necessarily estimate the treatment effect relevant to overall public health. There are instances, especially with vaccines and therapies for infectious diseases, that the appropriate effectiveness measures should include a broader focus on the overall public health. • The intent for the transition table and/or tool needs to be more explicit, whether intended for investigators to guide study design at protocol development, for reviewers to assess study proposals, and/or as criteria for evaluating study results. • While a framework is useful for general guidance, the optimal study design and analytic approach for a specific question is highly case specific.
19	Customer input and outcomes measures	• Encourage the use of advance tools such as QR codes, Smart Phone, and mobile devices to obtain the "voice of customer" easily and continuously. • Consider tools such as the Six Sigma Quality Function Deployment (QFD) to assist in prioritizing multiple options against multiple customer views. • Ensure outcome measures are aligned to the perspectives of patients and decision makers. • Drive common end points/outcomes/parameters on how treatments are measured. • Take the lead in simple reporting for patients and other decision makers to quickly uncover the pros and cons of different choices. • Allow and encourage the use of imperfect data in ways that best support decision-making.
20	Patient perspective	• Consider the study burden for study participants under ethical dimensions. • Include quality of life measures when the end of life is eminent. • If patients will be reading the definitions in Box 1, the language is far too complex. • In the wording "benefits and harms," harms is disproportionately negative and could be substituted with shortcomings, drawbacks, or disadvantages.
21	Cost	• Research of outcomes should include reporting cost information specific to the subject of the research – cost to patients and families, cost to institutions, cost to private insurers, cost to government programs.

Response	Focus	Comments
22	Case study	Case study: Percentage of patients, doctors, nurses and administrators knowledge of Ghana Patients Charter implementation in public clinic in Accra, Ghana.
23	Translation table	• The diagram lacks information that characterizes the causal pathway between treatment and outcome. • In conducting a time series analysis, the length of the time series used in models is very important and caution should be exercised regarding the certainty of statistical estimates whenever the time span is too short. • We are shifting from use of health care claims data to HER data for our time series and causal modeling. As payers shift from a fee-for-service model to a risk model, providers will have the incentive to accumulate more information to better understand how to manage patients, particularly those with chronic conditions and multiple comorbidities. • The conduct of randomized clinical trials is very time consuming and expensive. When RCTs are completed several unanswered questions remain about limitations of the medical intervention in the real world setting. Pragmatic clinical trials will address many of these unanswered questions regarding patient management and outcomes. • Marginal structural model (MSM) and inverse probability weighting (IPW) can be used to accommodate a wide variety of information sources outside of traditional health care sources such as socio-economic, workplace, etc. Because this type of analysis is longitudinal, the model will adjust for non-traditional factors such as patient preferences as well as health care factors. As electronic health record (EHR) information becomes available, less and less of the potential confounding will exist, thereby giving greater assurance of a casual interpretation of the findings and improvement in how a particular treatment will affect patient concerns and choices. Case study: Erythropoietin Stimulating Agents (ESA) for treatment of end-stage renal disease (ESRD) related anemia.
24	Case study	Case study: Women's Health Initiative Genome Education Program.

Appendix D-3
SUMMARY OF RFI CASE EXAMPLES

Response	Case study	Description
1	What is the comparative value of major drug-eluting stents in a high-risk, sub-population of diabetic patients?	*Possible Research Domains:* 1) Studies on the effectiveness of devices and; 2) Modeling the effectiveness of therapies for populations beyond primary evidence. *Decision-Maker Perspectives:* Focuses on government payer but has broad application to clinical and non-clinical stakeholders as well as patients. *Case Study Objective:* This case-study is intended to illustrate: 1) The limitation of small sample sizes for subgroup populations in RCTs of drug-eluting stents; 2) Absence of RCTs which directly compare all relevant drug-eluting stents competitors head-to-head; 3) Use of practical, methodological designs that extend beyond the RCT (i.e., meta-analysis and indirect treatment comparison (ITC)) to address the comparative value of drug-eluting stents.
2	Patient-centered preoperative empowerment education: a cost-effective intervention that improves quality of care and decreases length of stay	*Purposes:* To increase patients' self-efficacy scores of preoperative preparation, maximize patients' sense of empowerment, ensure patients' completion of preoperative preparation as outpatients, and decrease patients' length of hospital stay. *Importance:* An increasing number of elective surgery patients are being admitted on the day of surgery and have shorter lengths of hospital stay. Providing patients with sufficient perioperative knowledge and education within a confined time frame has been a challenge. The consequences of inadequate preoperative preparation and education can be costly and unsafe to the patient and the institution. Many studies have shown that patient specific preoperative education may empower patients with sufficient knowledge to enable them to carry out preoperative preparation completely and allow them to be an integral part of their own perioperative management. *Translational Framework:* "Knowledge to Action Process Framework" was adapted to translate evidence to clinical practice. *Conclusion:* A total of 122 adult urological cancer patients participated in the quality improvement project at NYU Preadmission Testing Unit, between January and March 2011. The study supported that a patient-centered preoperative empowerment intervention is a cost-effective strategy that improves quality of care and decreases length of stay.

Response	Case study	Description
4	Treatment choice decision tool for patients with CML	• To address a patient perspective in CML, we would define the target population through ICD-9 codes; validate them in our tumor registry, conduct a descriptive comparison of demographics, treatment, and outcomes assessment; and consider a QoL survey of patients based on a literature search on best possible surveys to date. We could also compare treatment and outcomes assessments against NCCN guidelines. All of this information could be incorporated into a decision tool for patients to assist them and their caregivers in understanding the impact of treatments, adverse events, outcomes, and cost to the patient and/or insurance.
4	How does a comprehensive medication review affect older adult medication use and outcomes?	This question is relevant to the expansion of Medicare Part D and the prioritization by CMS of CMRs as a key service within MTM Programs. Depending on priorities of payers vs. patients vs. providers, the question could be focused on particular patient subgroups. For example, payers might want evidence about the effects of CMRs on the most costly patients, while patients may prefer a study of patients with the greatest health or medication risks. Providers might want a study to address a patient subgroup most lacking in evidence to guide treatment choices. In a similar way, variation in preferences could influence study design decisions. For example, using only Medicare claims data retrospectively probably would limit the dependent variables to drug claims-based medication use measures and cost variables. This might fit best for payers' interests. Another retrospective design might complement claims data with patient surveys or patient record extractions, which could provide variables of interest to patients and providers (e.g. HRQoL, functional status, lab test results). It is possible that an RCT design could be utilized, though likely at greater cost and length of time, which might vary across perspectives.

Response	Case study	Description
4	What is the effect of continuing versus discontinuing medications for osteoporosis in bedridden patients with limited life expectancy on hospitalizations for fractures and health-related quality of life?	A recent systematic review noted the lack of available evidence to support decisions about continuing versus discontinuing calcium, vitamin D, and bisphosphonates in frail patients. The usual advantages of a well-designed RCT would apply to this question, namely high internal validity and the ability to very precisely measure relevant variables. However, such a trial may be limited to small numbers of patients in a specific setting (e.g., a single nursing home unit), decreasing the ability to precisely measure relationships between treatments and outcomes or explore heterogeneity of the effects in patient subgroups, and limiting generalizability to patients in other settings (e.g., community-dwelling patients receiving home care, or those not willing to be randomized to medication discontinuation). A RCT may also pose ethical concerns and high burden for patients and would likely be resource intensive. On the other hand, an analysis of existing administrative health care data tends to produce faster, more generalizable findings, can often be accomplished inexpensively, can include very large numbers of participants and thus more precise estimates of treatment-outcome relationship and exploration of treatment heterogeneity. These advantages are often at least partially offset by greater threats to internal validity from unmeasured differences between treatment groups and difficulty in precisely measuring relevant variables. However, there exists an increasing array of rigorous statistical methods to address confounding in observational treatment-outcome studies that often have the ability to thoroughly address these concerns (e.g., instrumental variables, propensity scores, marginal structural models). It is impossible to judge the amount of bias introduced in an observational study without understanding the methodological details and analysis approach. Similarly, if the study requires linking multiple complex databases from multiple sources and stakeholders (e.g., claims “owned” by the payer to health record data owned by multiple health systems to survey data collected by researchers), the resources and infrastructure required may be as extensive as that for an RCT, and logistical issues with regard to HIPAA and date use agreements may become problematic.
4	How will a proposed therapy/drug impact my disease (an oncology perspective)?	Patients often ask how a proposed therapy/drug will impact their diseases. The response is generally a generic approach from a recent clinical trial. A better approach would provide information to help patients understand how specific “patients like them” responded to that therapy. Outcomes would likely vary between patients, such as a 60 year old in poor health with many comorbidities compared to a 47 year old in relatively good health. Yet, the generic approach uses the same assumptions of probability of benefit with treatment for all patients. Patient-centered research would help refine the evidence and information available to patients to help them make better decisions about their care options.

Response	Case study	Description
5	Studies of care transitions	The early insights on care transitions came from RCTs. However, in situations in which services are disastrously error prone, virtually any reasonable RCT will show improvement. However, the RCT design prohibits use of some of the strongest system redesign opportunities because they contaminate the control group. Alternative study designs with control charts, workflow innovations and role redesign, and rapid feedback in a variety of settings would quickly provide substantial insight to improve patient-centered care.
6	Comparative effectiveness of treatments of TMJ disorders	As in any orofacial pain management, the dentist is an integral part of a pain group that involves pain physicians, physical therapists, mental health, and rehabilitation specialists. The multidisciplinary approach gives the patient a variety of treatment modalities: home care, physical therapy, intraoral appliances, pharmacology, neurology and psychiatry. The perpetuation of TMJ disorders revolves around a constellation of factors—structural and neurological—and the practitioner should strive for a balanced approach that addresses both. A balance between local and systemic intervention is also important. Compressions, torques and deviations of the mandible can all be addressed by working on the local muscles and fasciae. Once balanced on a local level, practitioner may try integrating that balance with larger patterns in the rest of the body. Doing this work can make the benefits more profound and long-lasting for the patients. All patients vary with regard to their pathology and clinical characteristic, leading the clinician to adopt a multi-disciplinary symptom-oriented approach to care.
6	Comparative effectiveness of pharmacological management of alcoholic liver disease	Treatment modalities for alcoholic liver disease (ALD) and acute alcoholic hepatitis (AAH) are insufficient. To date, none of the therapies proposed has been shown to consistently improve the course of alcoholic liver damage, and there is no FDA approved therapy for ALD. Some drugs, in particular corticosteroid and pentoxifylline, appear to be beneficial in the subgroup of patients with severe AAH, while others, such as insulin–glucagon (pro-growth), polyunsaturated phosphatidylcholine (antioxidant, TNF-a modulator, anti-fibrotic), and vitamin E, have been proposed for the treatment of ALD, but none of them have shown a convincing benefit.
6	Comparative effectiveness of HAART in HIV-infected patients with cardiac disease	Patients with AIDS can have cardiac pathology related to opportunistic infections and tumors as well as that related to antiviral medication (highly active antiretroviral therapy, HAART). The frequency of cardiac manifestations is influenced by different variables including survival prolongation in HIV infected patients because of advances in antiretroviral treatment and reduction in the occurrence of opportunistic infections. With the introduction of highly active antiretroviral therapy (HAART), including protease inhibitor, patients are living longer and other co-morbidities such as hypertension; metabolic abnormalities including hyperglycemia and hyperlipidemia; and lipodystrophy increase their risks of cardiovascular diseases.

Response	Case study	Description
6	Comparative effectiveness of management of mandibular condyle fractures	There are three main unresolved issues in the management of condylar fractures: 1) the method of treatment (open vs. closed and age); 2) approach to open treatment (extra-oral, intraoral, or endoscopic); and 3) fixation vs. no fixation and the types of fixation. The conclusive evidence based on collective reports of experiences documented in the literature includes: 1) intracapsular fractures are best treated closed; 2) fractures in children are best treated closed except when the fracture itself anatomically prohibits jaw function; 3) most fractures in adults can be treated closed except in cases of gross displacements of the fragments and/or severe dislocations of the condylar head; 4) physical therapy that is goal-directed and specific to each patient is integral to good patient care and is the primary factor influencing successful outcomes, whether the patient is treated open or closed; 5) when open reduction is indicated, the procedure must be performed well, with an appreciation for the patient's occlusal relationships, and it must be supported by an appropriate physical therapy and follow-up regimen; and 6) fractures that are openly reduced can be rigidly or semi-rigidly fixed in very stable physiologic position.
11	Delphi survey of research question and study design preferences	To "ground" the methods recommendations in real-world contexts, we commissioned two teams of methodologists to write white papers on appropriate study design considerations for four specific PCOR case studies. These case studies were then tested with a 21-person stakeholder panel comprised of patients/consumers, providers, methodologists, government agencies, public and private payers, industry representatives, clinical guideline developers, and technology assessment organizations. Stakeholders completed pre-meeting surveys rating the potential value of four study designs and participated in pre-meeting individual interviews to clarify their underlying rationale for their responses. At the in-person meeting, aggregate ratings from the pre-meeting survey were shared along with their individualized ratings. Stakeholders discussed each case study and completed a second survey to elicit their preferences for various attributes and study designs. At the conclusion of the meeting, feedback was solicited to understand general factors that may guide study selection beyond the three selected case studies.
12	Comparative effectiveness of surgery and radiation for localized prostate cancer focusing on modern technologies and controlling for differences in patients and treatments that may affect outcomes	While clinical trials with random allocation of radical prostatectomy and radiation therapy would be the ideal study design to assess the relative effectiveness for survival and health-related quality of life, this study design is impractical due to the long follow-up and large number of patients needed. One approach to assess the comparative effectiveness of surgery and radiation for localized prostate cancer is to use retrospective observational studies for morbidity and mortality and RCTs for quality of life. To allow for the appropriate inferences from observational studies using administrative data, it is necessary to make the appropriate adjustments for bias and confounding. By doing so, and although their confidence in making evidence-based decisions would remain rather low, observational studies may provide the best evidence for the stakeholders in ascertaining the extent to which the available contemporary treatments differ in their effectiveness on patient important outcomes.

Response	Case study	Description
12	Compare the effectiveness of comprehensive care coordination programs, such as the medical home, and usual care in managing children and adults with severe chronic disease, especially in populations with known health disparities	Medical homes or Accountable Care Organizations are currently being advanced as a recent health care policy initiative to facilitate comprehensive care coordination, improve patient-centered outcomes, and reduce health care costs and unnecessary resource utilization. To answer a comparative effectiveness research question for medical homes, a useful approach would be a pragmatic clustered randomized trial among appropriately selected patients (multiple chronic illnesses requiring multi-disciplinary care) with clearly defined outcomes (patient-centered outcomes, health care costs, and resource utilization). This research design to assess the effectiveness of medical homes would be guided by the costs associated with creating a medical home and buy-in from primary care providers and specialists. A pragmatic clustered randomized trial to ascertain the effectiveness of medical homes would provide the most generalizable and applicable estimates of treatment effect.
22	Percentage of patients, doctors, nurses, and administrators with knowledge of the Ghana Patients Charter implementation in public clinics in Accra, Ghana	The research revealed that not all patients were aware of their patients' rights, but all administrators, nurses, and doctors were well acquainted with the policy at all the clinics visited. For the policy to be well implemented, there needs be awareness at all public and private clinics to educate patients about their rights. This can best be achieved by equipping both local and international volunteers in conjunction with visionary and proactive NGOs, like Concern Health, working with Commission of Human Rights Administrative Justice in Ghana to ensure that the Patient's Charter is well implemented throughout Ghana, not just in Accra.

Response	Case study	Description
23	Erythropoietin Stimulating Agents (ESA) for treatment of end-stage renal disease (ESRD)	To determine whether FDA's 2007 ESA safety warning had an impact on the patient-clinician decisions that affected ESA practice patterns, Thamer et al, using a time series model, analyzed ESA prescribing patterns for all U.S. adult hemodialysis patients in the year before (2006) and nearly two years (2007-8) after an FDA black box warning *to use the lowest possible ESA dose*. They found that the average ESA dose/administration in the follow-up period was significantly negative. The declines in average ESA dose were greatest for anemic patients and those maintained within the FDA-recommended hematocrit target range. There was also a large decline in dose among patients at the second largest for profit dialysis chain. On average, they found that ESA doses were reduced by 12 units/administration per month in the nearly two years that followed the March 2007 FDA black box warning. They concluded that nephrologists and dialysis providers did heed the FDA black box warning to reduce ESA doses, albeit modestly, for their dialysis patients. The extent to which patient-physician communication improved regarding the decision to use ESA therapy and the appropriate dose is unknown. Since 2007, FDA has issued another decision to reduce the upper bound of the ESA therapy target from a hemoglobin level of 12 g/dL to 11 g/dL effective as of June 2011. As of January 2011, CMS has changed its reimbursement policy from a fee-for-service to a fixed payment scheme (referred to ESRD PPS). Early indications suggest that both ESA use and hematocrit levels have been dramatically reduced since implementation of ESRD PPS. As data become available, we propose to examine if ESRD PPS has resulted in a further (and perhaps more dramatic) decline in ESA doses compared to both the 2007 and 2011 the FDA black box warnings and whether these policies changes translated to informed decision-making that incorporated the best health care knowledge into the application of care and patient outcomes. The key role played by patients in the changes in ESA therapy need to be examined and quantified.
24	Women's Health Initiative Genome Education Program	While genome-wide association studies (GWAS) provide an opportunity to identify genetic polymorphisms that may be associated with certain disease conditions, with the goal of improving health and health outcomes, genetic studies in general have the potential to increase concern about the impact on their families, racial and ethnic differences, ability to get insurance, and impact on employment. While this is true in general, it is increased among minority populations. One might anticipate that this also may be true in older populations because of their history and lack of exposure to genetics in general during their lifetimes. The purpose of this project is to develop an educational tool to inform the women in the Women's Health Initiative GWAS of the implications and knowledge to be gained from the studies and to allay any concerns they may have for themselves and their families.

Appendix D-4

TRANSLATION FRAMEWORK CASE STUDIES

Three case studies illustrate the specification of a research question and the thinking behind making appropriate study design, data source, and analytic choices using the translation framework approach.

Example 1: High-intensity statin mono-therapy versus fixed simvastatin-ezetimibe combination therapy

Develop the Research Question

In primary prevention of coronary heart disease, will simvastatin-ezetimibe combination therapy use lead to better outcomes compared to statin only treatment?

Patients: Asymptomatic adults age $\geq$ 35 years in men and $\geq$ 45 years in women. No conditions that limit usual levels of physical activity (i.e. able to walk at least 2 flights of stairs without dyspnea or chest discomfort). Low-intermediate background risk of coronary heart disease (Framingham risk score 6-10); no symptoms or signs of peripheral vascular or carotid disease; no family history of premature coronary disease (MI or sudden death in first degree male relative age $\leq$ 45 years, female relative age $\leq$ 55 years); and no illnesses associated with a life expectancy < 2 years (e.g. severe COPD, liver dysfunction, metastatic cancer).

Intervention: New users of Vytorin (simvastatin + ezetimibe in fixed combination).

Comparator: New users of high intensity statin therapy without ezetimibe.

Generic name	Brand name	High-intensity daily doses (mg)
Atorvastatin	Lipitor	> 10
Lovastatin	Mevacor	> 40
Rosuvastatin	Crestor	> 5
Simvastatin	Zocor	> 40

Outcomes: Event rates over time including, mortality, myocardial infarction, hospitalization for unstable angina that includes unplanned revascularization. Quality of life measures over time and other patient-reported outcomes.

Timing and setting: Primary care physicians in outpatient (non-emergency room) practices with patients with stable health care access and insurance with prescription drug coverage for statins and Vytorin.

Prior evidence: Survey data on existing practices and attitudes indicate that Vytorin is frequently used in the US, much more so than in Canada. Evidence supports its efficacy against placebo. The ENHANCE trial however, could not demonstrate meaningful changes in measured arterial intima thickness when compared to simvastatin in patients with familial hypercholesterolemia and was not powered to detect differences in cardiovascular events. IMPROVE-IT is a large randomized trial that is ongoing and recruitment has been extended.

Intent of research, decisions to be made, and stakeholder perspectives: Determined by the need that the research will address.

Determine the Type of Evidence Needed by Defining Study Characteristics

Resources and feasibility: Consider population-based cohorts (Nurses', Health Professionals). EHR data is available within some integrated health care systems (e.g. HMO). Research networks might be willing and/or able to do large-scale database studies: FDA Mini Sentinel, HMO RN, VA, DOD, commercial insurance plans. Consider a de-novo randomized trial and registries. Review guidance documents, criteria for primary prevention of CHD, and treatment guidelines for hypercholesterolemia.

Determine Research Category

Research Category: Effectiveness of therapeutics

Navigate the Tradeoffs between Various Methodologic Choices

Study Design and Analysis

- Large-scale randomized trial with active comparator and focus on "hard" clinical endpoints.

Advantages include: balanced treatment groups at baseline; standard endpoint assessment and adjudication when necessary; credible results that might be more readily accepted by stakeholders.

Disadvantages include: high cost; difficulty recruiting participants; long time until an answer is obtained; event rates may be lower than predicted; and risk of cross-contamination.

- Small-scale randomized trial with surrogate endpoints

Advantages include: lower cost; balanced treatment groups at baseline; standard endpoint assessment and adjudication when necessary.

Disadvantages include: study will be underpowered for hard clinical endpoints; results may be of limited use for clinical practice; and risk of cross-contamination. Also, this type of trial has already been done to address the research question (ENHANCE trial).

- Large-scale cluster randomized trial

Advantages include: reduced risk of cross-contamination; observer bias may be less severe; the trial may be closer to "real life" because it would reflect practice or systems policies; balanced treatment groups at baseline; standard endpoint assessment and adjudication when necessary; credible results that might be more readily accepted by stakeholders.

Disadvantages include: high cost with large sample size requirements; long time until an answer is obtained; cross-over may be larger than anticipated; event rates may be lower than predicted.

- Adaptive trial

Advantages include: may require smaller sample size; may be completed in less time; balanced treatment groups at baseline; credible results that might be more readily accepted by stakeholders.

Disadvantages include: difficult to budget and plan; long time until an answer is obtained; event rates may be lower than predicted; and risk of cross-contamination.

- Observational cohort study in secondary health care databases with application of propensity score matching and/or instrumental variables

Advantages include: lower cost; short duration; reflects patients and treatment choices outside of controlled research environments; will likely yield high precision; and may allow multiple subgroup analyses.

Disadvantages include: patient groups may not be balanced by unobserved characteristics; failure to meet instrumental variable assumptions; and risk of cross-contamination.

Data Sources

- De-novo data

Advantages include: standardized assessment of exposure, outcomes, and patient characteristics; completeness of patient assessment at pre-defined time points (e.g. baseline, 6 months, etc.)

Disadvantages include: high costs; long time for data collection; and selective groups of participants.

- Electronic Health Records within integrated health care systems

Advantages include: passive acquisition of data; large numbers; lower cost; possibly reduced observer bias; and representative population.

Disadvantages include: coding inaccuracies and misclassification, especially of endpoints; and higher rates of loss to follow-up.

- Health insurance claims data

Advantages include: passive acquisition of data; very large numbers; lower cost; possibly reduced observer bias; and representative population.

Disadvantages include: incomplete capture of patient health state and "softer" endpoints; coding inaccuracies and misclassification, especially of endpoints; and higher rates of loss to follow-up.

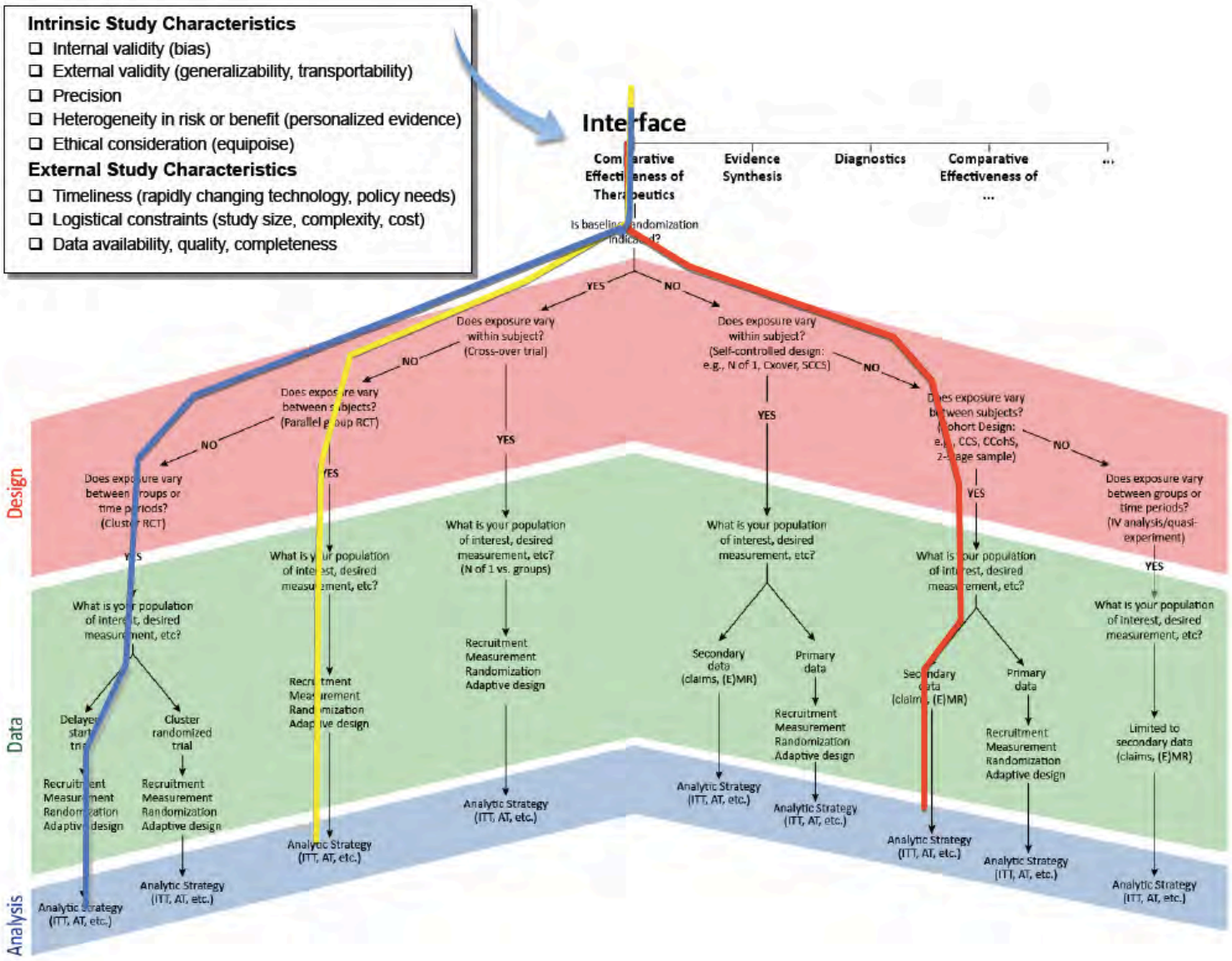
- Registries (previously collected data)

Advantages include: passive acquisition of data; possibly reduced observer bias; and lower risk of coding inaccuracies and misclassification.

Disadvantages include: higher cost; missing data; and follow-up of varying quality.

These considerations lead to three preferred pathways through the translation framework, illustrated as a branch diagram in the figure below (see pathways in red, blue, and yellow). Study characteristics would be prioritized differently depending on the evidentiary needs of the stakeholders. A non-randomized study is depicted by the red path. This choice assumes that confounding caused by treatment selection according to baseline risk could be controlled through design and analytic choices. By using secondary data, this choice is also compatible with the desire to get results expeditiously and to more closely represent health care practice.

A randomized study would generally be preferred when validity is a high priority (yellow and blue paths). To provide substantial levels of control of the research environment, a parallel group randomized trial with patient-level randomization could be considered (yellow path). To reduce the impact of the study on health care delivery, a cluster randomized trial that makes either treatment exclusively available to entire population groups could be an option (blue path). This design could also increase generalizability, which may be a priority to stakeholders.


Intrinsic Study Characteristics
Internal validity (bias)
External validity (generalizability, transportability)
Precision
Heterogeneity in risk or benefit (personalized evidence)
Ethical consideration (equipoise)
External Study Characteristics
Timeliness (rapidly changing technology, policy needs)
Logistical constraints (study size, complexity, cost)
Data availability, quality, completeness
Interface
Comparative Effectiveness of Therapeutics
Evidence Synthesis
Diagnostics
Comparative Effectiveness of
...
Is baseline randomization indicated?
YES
NO
Does exposure vary within subject? (Cross-over trial)
Does exposure vary between subjects? (Parallel group RCT)
Does exposure vary between groups or time periods? (Cluster RCT)
Does exposure vary within subject? (Self-controlled design: e.g., N of 1, Cxover, SCCS)
Does exposure vary between subjects? (Cohort Design: e.g., CCS, CCohS, 2-stage sample)
Does exposure vary between groups or time periods? (IV analysis/quasi-experiment)
What is your population of interest, desired measurement, etc?
What is your population of interest, desired measurement, etc? (N of 1 vs. groups)
Delayed start trial
Cluster randomized trial
Recruitment Measurement Randomization Adaptive design
Secondary data (claims, (E)MR)
Primary data
Limited to secondary data (claims, (E)MR)
Analytic Strategy (ITT, AT, etc.)
Design
Data
Analysis

Example 2: Coronary screening in patients with asymptomatic coronary heart disease

Develop the Research Question

- In asymptomatic adults with Framingham risk score 6-10, will measurement of coronary artery calcium lead to better outcomes?

- Patients: Asymptomatic adults, age $\geq$ 35 years in men and $\geq$ 45 years in women, with no condition that limits usual levels of physical activity (i.e. able to walk at least 2 flights of stairs without dyspnea or chest discomfort.) Low-intermediate background risk of coronary heart disease (Framingham risk score 6-10); no symptoms or signs of peripheral vascular or carotid disease; and no family history of premature coronary disease (MI or sudden death in first degree male relative age $\leq$ 45 years, female relative age $\leq$ 55 years). No diabetes; estimated creatinine clearance > 60; no illnesses associated with a life expectancy < 2 years (e.g. severe COPD, liver dysfunction, metastatic cancer).

- Intervention: Measurement of coronary artery calcium (CAC) deposition by CT (chosen because of its ability to reclassify risk) and prescription of simvastatin 40 mg po qd if CAC $\geq$ 100 irrespective of LDL cholesterol.

- Comparator: Calculation of Framingham risk score alone with no measurement of CAC, and prescription of simvastatin 40 mg po qd if ATP III criteria are met (should be rare in asymptomatic adults with Framingham risk score of 6-10).

- Outcomes: Event rates over time including mortality, myocardial infarction, hospitalization for unstable angina that includes unplanned revascularization, and major complications of interventional procedures. Other major adverse clinical events, quality of life measures over time, patient-reported outcomes, and personal values and preferences.

- Timing and setting: Primary care physicians in outpatient (non-emergency room) practices with patients with stable health care access and insurance. There are no state laws addressing CAC (e.g. not Texas) and no insurance coverage of CAC, but other preventive services are covered including prescription drug coverage for simvastatin. Assume that time is available to conduct a study that could provide a definitive answer and CAC technology is stable.

- Prior evidence: Survey data on existing practices and attitudes informs the value of information and focuses on patient-centered variables. This evidence has been reviewed by USPSTF, ACC, and AHA. Cohort studies showing that CAC predicts coronary events and reclassifies risk are available. A few small underpowered randomized trials with surrogate endpoints have been published, but no large scale randomized trials analogous to National Lung Screening Trial.

Determine the Type of Evidence Needed by Defining Study Characteristics

- Resources and feasibility: Population-based cohorts with routine measurement of CAC, ACC/NCDR outpatient registry, EHR data available within some integrated health care systems (e.g. HMO), Research networks that might be willing and/or able to do large-scale randomized trial: HMO, VA, military, independent, Guidance documents, Criteria for valid screening tests in asymptomatic populations as articulated various bodies (e.g. USPSTF), Existing consensus on standards of care.

Determine Research Category

Effectiveness of diagnostic test and imaging strategies

Navigate the Tradeoffs between Various Methodologic Choices

Design and Analyses

- Large-scale randomized trial analogous to National Lung Screening Trial with focus on "hard" clinical endpoints, intent-to-treat

Advantages include: accounts for unmeasured confounders, lead-time, length-time, and overdiagnosis biases; and credible results that might be more readily accepted by stakeholders.

Disadvantages include: high cost; difficulty recruiting; long time until an answer is obtained; newer technologies and information may render CAC obsolete; event rates may be lower than predicted; risk of cross-contamination; and observer bias due to lack of blinding.

- Small-scale randomized trial analogous to DIAD with focus on "softer" clinical endpoints

Advantages include: lower cost; and for endpoints measured, accounts for lead-time, length-time, and overdiagnosis biases.

Disadvantages include: underpowered for hard clinical endpoints; risk of false positive findings for hard clinical endpoints; results may not be considered credible by stakeholders; newer technologies may render CAC obsolete; risk of cross-contamination; and observer bias due to lack of blinding.

- Large-scale cluster randomized trial

Advantages include: avoids cross-contamination; observer bias may be reduced; may be more representative of "real life" because it would reflect practice or systems policies; accounts for

unmeasured confounders, lead-time, length-time, and overdiagnosis biases; and results would be more credible for stakeholders.

Disadvantages include: high cost with high sample size requirements; long time until an answer is obtained; newer technologies may render CAC obsolete; and event rates may be lower than predicted.

- Adaptive trial

Advantages include: may require a smaller sample size and could be completed in less time; accounts for unmeasured confounders and lead-time, length-time, and overdiagnoses biases; and results would be more credible for stakeholders.

Disadvantages include: difficult to budget and plan; long time until an answer is obtained; newer technologies and information may render CAC obsolete; event rates may be lower than predicted; risk of cross-contamination; and observer bias due to lack of blinding.

- Observational cohort study with application of propensity score matching and/or instrumental variables

Advantages include: lower cost; "real life" populations depending on the data source; and may yield acceptable effect-size estimates.

Disadvantages include: may not account for unmeasured confounders; failure to meet instrumental variable assumptions; may not account for lead-time, length-time, and overdiagnosis biases; and observer bias due to lack of blinding.

- Data Sources

- EHR within integrated health care systems

Advantages include: passive acquisition of data; lower cost; and possibly reduced observer bias.

Disadvantages include: non-representative populations; coding inaccuracies and misclassification, especially of endpoints; and higher rates of loss to follow-up.

- Registries (e.g. NCDR)

Advantages include: passive acquisition of data; possibly reduced observer bias; and lower risk of coding inaccuracies and misclassification.

Disadvantages include: non-representative populations; selection bias; and higher cost.

Example 3: Drug-eluting stent for stable angina due to moderate severity coronary artery disease

Develop the Research Question

- In adults with medication-resistant stable angina pectoris that is caused by single-vessel mid-LAD disease, will placement of a drug-eluting stent, as opposed to a bare metal stent, lead to better short- and long-term patient-centered outcomes?

- Patients: Adults with stable angina pectoris, age $\geq$ 30 years; other than angina, no conditions that limit usual levels of physical activity (e.g. except for angina, would be able to climb at least two flights of stairs). Angina not adequately controlled with optimal medical therapy. Patients have hemodynamically significant single-vessel disease involving the mid LAD (blockage > 70% severity, not a bifurcation lesion) and documented myocardial ischemia by exercise ECG, stress SPECT, or stress echocardiography. The lesion meets FDA "on-label" criteria for drug-eluting stents and the lesion can be successfully treated with one stent (i.e. no need for overlapping or multiple stents). Comorbidities include diabetes allowed, estimated creatinine clearance > 60, but no illnesses associated with a life expectancy < 2 years (e.g. severe COPD, liver dysfunction, metastatic cancer). Patients should have no contraindications to long-term dual anti-platelet therapy and no anticipated surgeries in the next year.

- Intervention: Drug-eluting stent (paclitaxel, sirolimus, or "next generation") and at least one year of dual anti-platelet therapy.

- Comparator: Bare metal stent and 3 months of dual anti-platelet therapy followed by aspirin mono-therapy.

- Outcomes: Event rates over time including mortality, myocardial infarction, and hospitalization for unstable angina that includes unplanned revascularization; target lesion revascularization; major complications of interventional procedures; major bleeding leading to transfusion or hemodynamic collapse; other major adverse clinical occurrences; quality of life measures over time; additional patient-reported outcomes and personal values and preferences; benefit of lower chance of repeat procedure; and risks and discomforts associated with long-term dual anti-platelet therapy.

- Timing and setting: Referral to interventional cardiologist for management of medication-resistant angina; hospital with interventional capabilities; interventional cardiologists with typical levels of experience and complication rates; stable health care access and insurance coverage for procedures and for long-term clopidogrel or prasugrel therapy; available time to establish a definitive answer; and stable stent technology.

- Prior evidence: Survey data on existing practice and attitudes (e.g. NCDR) that inform value of information and focus on patient-centered variables; evidence reviews by ACC/AHA, Cochrane collaboration and others. Multiple randomized trials have been completed that indicate drug-eluting stents reduce the need for target-vessel or target-lesion revascularization, although there is no clear reduction of death or myocardial infarction. No single large-scale trials address risk of death or myocardial infarction and no trials have addressed quality of life or functional capacity.

Determine the Type of Evidence Needed by Defining Study Characteristics

- Resources and feasibility: ACC/NCDR registries; AHA Get-with-the-Guidelines; EHR data available within integrated health care systems (e.g. HMO); research networks that might be willing to do large-scale trials that address QOL: HMO, VA, military, independent; guidance documents; ACC/AHA guidelines and appropriateness statements; and existing consensus on standards of care.

Determine Research Category

Research Category: Effectiveness of implantable devices

Navigate the Tradeoffs between Various Methodologic Choices

- Design and Analyses
- Large-scale randomized trial
- Consider both superiority and non-inferiority designs. Superiority may be more appropriate for Quality of Life measures and non-inferiority may be more appropriate for hard clinical events other than target vessel- or target lesion revascularization.

Advantages include: accounts for unmeasured confounders; allow for less biased ascertainment of quality of life; credible results that might be more readily accepted by stakeholders; economic pressures and realities might generate enthusiasm for a large-scale patient-centered trial.

Disadvantages include: for "hard clinical endpoints" would expect a small effect size and would therefore need very large samples; difficulty recruiting in an environment that largely favors drug-eluting stents; high cost; quality of life endpoints are difficult to translate into real-life clinical decisions; newer stent technologies may render results obsolete; rates of hard clinical events will be

low; observer bias because blinding may be difficult to accomplish; and long-time to obtain an answer.

- Smaller-scale randomized trial with focus on quality of life

Advantages include: accounts for unmeasured confounders; economic pressures and realities might generate enthusiasm; and results may be considered credible by patients.

Disadvantages include: results may not be considered credible by some stakeholders, because QOL endpoints may be difficult to translate into real-life clinical decisions; newer stent technologies may render results obsolete; observer bias because blinding may be difficult to accomplish; moderately long time to obtain an answer.

- Adaptive trial

Advantages include: may require smaller sample size and be completed in less time; account for unmeasured confounders; credible results, particularly for hard clinical events; and economic pressures may generate enthusiasm for a trial with a design that may produce a quicker, yet robust, answer.

Disadvantages include: difficult to plan and budget; newer technologies may render findings obsolete; and low event rates.

- Observational cohort study with application of propensity score matching, instrumental variables, or before-after design

Advantages include: lower cost; "real life" populations depending on data sources; and may yield acceptable effect-size estimates if samples are large enough.

Disadvantages include: may not account for unmeasured confounders; failure to meet instrumental variable assumptions; observer bias due to lack of blinding; and quality of life data may not be available or reliable.

- Data Sources
- EHR within integrated health care systems

Advantages include: passive acquisition of data and lower cost.

Disadvantages include: non-representative populations; coding inaccuracies and misclassification; higher rates of loss to follow-up; and absence of standardized, validated QOL data.

- Society or government registries (e.g. NCDR)

Advantages include: passive acquisition of data; standardized definitions with lower risk of coding inaccuracies and misclassification; and may be possible to collect standardized, validated QOL data.

Disadvantages include: non-representative populations; selection bias; absence of blinding; higher cost; and acquisition of standardized, validated QOL data may impose undue burden on data collection effort.

Appendix E: Relevant Sections of the Patient Protection and Affordable Care Act

E-1: Definition of comparative clinical effectiveness research

"(A) IN GENERAL.— The terms 'comparative clinical effectiveness research' and 'research' mean research evaluating and comparing health outcomes and the clinical effectiveness, risks, and benefits of 2 or more medical treatments, services, and items described in subparagraph (B).
(B) **MEDICAL TREATMENTS, SERVICES, AND ITEMS DESCRIBED**.—The medical treatments, services, and items described in this subparagraph are health care interventions, protocols for treatment, care management, and delivery, procedures, medical devices, diagnostic tools, pharmaceuticals (including drugs and biologicals), integrative health practices, and any other strategies or items being used in the treatment, management, and diagnosis of, or prevention of illness or injury in, individuals."
124 STAT 727

E-2: Purpose of the Institute

"The purpose of the Institute is to assist patients, clinicians, purchasers, and policy-makers in making informed health decisions by advancing the quality and relevance of evidence concerning the manner in which diseases, disorders, and other health conditions can effectively and appropriately be prevented, diagnosed, treated, monitored, and managed through research and evidence synthesis that considers variations in patient subpopulations, and the dissemination of research findings with respect to the relative health outcomes, clinical effectiveness, and appropriateness of the medical treatments, services, and items described in subsection (a)(2)(B)."
124 STAT 728

E-3: Duties of the Institute

"(A) **IDENTIFYING RESEARCH PRIORITIES**.—"The Institute shall identify national priorities for research, taking into account factors of disease incidence, prevalence, and burden in the United States (with emphasis on chronic conditions), variations and health disparities in terms of delivery and outcomes of care, the potential for new evidence to improve patient health, well-being, and the quality of care, the effect on national expenditures associated with a health care treatment, strategy, or health conditions, as well as patient needs, outcomes, and preferences, the relevance to patients and clinicians in making informed health decisions, and priorities in the National Strategy for quality care established under section 399H of the Public Health Service Act that are consistent with this section.
(B) **ESTABLISHING RESEARCH PROJECT AGENDA.—** The Institute shall establish and update a research project agenda for research to address the priorities identified under subparagraph (A), taking into consideration the types of research that might address each priority and the relative value (determined based on the cost of conducting research compared to the potential usefulness of

the information produced by research) associated with the different types of research, and such other factors as the Institute determines appropriate."
(2) **CARRYING OUT RESEARCH PROJECT AGENDA**.—
(A) RESEARCH.— The Institute shall carry out the research project agenda established under paragraph (1)(B) in accordance with the methodological standards adopted under paragraph (9) using methods, including the following:
(i) Systematic reviews and assessments of existing and future research and evidence including original research conducted subsequent to the date of the enactment of this section.
(ii) Primary research, such as randomized clinical trials, molecularly informed trials, and observational studies.
(iii) Any other methodologies recommended by the methodology committee established under paragraph (6) that are adopted by the Board under paragraph (9)."
124 STAT 728-9

E-4: Taking into account potential differences

"Research shall be designed, as appropriate, to take into account the potential for differences in the effectiveness of health care treatments, services, and items as used with various subpopulations, such as racial and ethnic minorities, women, age, and groups of individuals with different comorbidities, genetic and molecular sub-types, or quality of life preferences and include members of such subpopulations as subjects in the research as feasible and appropriate."
124 STAT 730-1

E-5: Functions of the Committee

"(C) FUNCTIONS.—Subject to subparagraph (D), the methodology committee shall work to develop and improve the science and methods of comparative clinical effectiveness research by, not later than 18 months after the establishment of the Institute, directly or through subcontract, developing and periodically updating the following:
(i) **Methodological standards for research.** Such methodological standards shall 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. Any methodological standards developed and updated under this subclause shall be scientifically based and include methods by which new information, data, or advances in technology are considered and incorporated into ongoing research projects by the Institute, as appropriate. The process for developing and updating such standards shall include input from relevant experts, stakeholders, and decisionmakers, and shall provide opportunities for public comment. Such standards shall also include methods by which patient subpopulations can be accounted for and evaluated in different types of research. As appropriate, such standards shall build on existing work on methodological standards for defined categories of health interventions and for each of the major categories of comparative clinical effectiveness research methods (determined as of the date of enactment of the Patient Protection and Affordable Care Act).
(ii) A **translation table** that is designed to provide guidance and act as a reference for the Board to determine research methods that are most likely to address each specific research question."
124 STAT 732-3

E-6: Reports to be submitted to the Board

"The methodology committee shall submit reports to the Board on the committee's performance of the functions described in subparagraph (C). Reports shall contain recommendations for the Institute to adopt methodological standards developed and updated by the methodology committee as well as other actions deemed necessary to comply with such methodological standards."
124 STAT 733

E-7: Release of research findings

"The Institute shall, not later than 90 days after the conduct or receipt of research findings under this part, make such research findings available to clinicians, patients, and the general public. The Institute shall ensure that the research findings—
(i) convey the findings of research in a manner that is comprehensible and useful to patients and providers in making health care decisions;
(ii) fully convey findings and discuss considerations specific to certain subpopulations, risk factors, and comorbidities, as appropriate;
(iii) include limitations of the research and what further research may be needed as appropriate;
(iv) not be construed as mandates for practice guidelines, coverage recommendations, payment, or policy recommendations; and
(v) not include any data which would violate the privacy of research participants or any confidentiality agreements made with respect to the use of data under this section."
124 STAT 733-4

Appendix F: Glossary

Terms	A number of words or phrases are used throughout this Report. Here are brief definitions for them.	Reference
A priori	From the Latin for "from the previous." Usually a hypothesis or decision made before a study begins.	1
Ad hoc	From the Latin for "for this." Often referring to a procedure or method selected in a specific circumstance without reference to a predetermined plan or scheme.	1
Adaptive clinical trial or study	A study that includes a prospectively planned opportunity for modification of one or more specified aspects of the study design and hypotheses based on analysis of data (usually interim data) from subjects in the study. Analyses of the accumulating study data are performed at prospectively planned timepoints within the study, can be performed in a fully blinded manner or in an unblinded manner, and can occur with or without formal statistical hypothesis testing. Adaptive designs use accumulating data to decide on how to modify certain aspects of a trial according to a pre-specified plan.	2,3
Adaptive randomization	Adaptive randomization is a form of treatment allocation in which the probability of patient assignment to any particular treatment group of the study is adjusted based on repeated comparative analyses of the accumulated outcome responses of patients previously enrolled (often called *outcome dependent randomization*, for example, the *play the winner* approach).	
Algorithm	An explicit description of an ordered sequence of steps with branching logic that can be applied under specific clinical circumstances. The logic of an algorithm is as follows: if a, then do x; if b, then do y; etc.	4
Automaton	Something that operates by following rules without needing human intervention or guidance.	1
Baseline	A condition, characteristic or circumstance existing or measured at the beginning of a study.	5
Bayesian Bayesian statistics:	A statistical approach that uses Bayes' Theorem to integrate related data and a-priori belief with observed data to estimate unknown parameters. This represents an alternative to the traditional frequentist approach that attempts to establish confidence intervals around parameters, and/or falsify *a-priori* null-hypotheses. Bayesian statistics is an approach for learning from evidence as it accumulates.	6, 7, 8
Bayesian or adaptive design trial	A trial that uses information that accumulates during the course of the trial as well as prior information to decide how to modify aspects of the trial as it continues.	

Terms	A number of words or phrases are used throughout this Report. Here are brief definitions for them.	Reference
Bias	A systematic tendency for the estimate of treatment effect to deviate from its true value A systematic (consistent, non-random) error that results in over or under-estimation of a parameter. Bias reflects the degree to which the statistic inaccurately measures the parameter that it is intended to estimate. Conflicts of interest can lead to biased design, conduct, analysis, and interpretation of study results.	2, 4, 6
Blinded analyses	Blinded analyses are those in which the treatment group assignments of study subjects are not known and are therefore not used in any manner in the analysis.	2
Blinding	Study procedure to keep specific information about the study unknown to investigators, participants, or both. Blinded analyses are those in which the treatment group assignments of study subjects are not known and are therefore not used in any manner in the analysis.	9
Case Control Study	Nonrandomized study of patients with a specific outcome (cases) compared to patients without the outcome (controls).	8
Causality	Relating causes to the effects they produce… Several types of causes can be distinguished. A cause is termed "necessary" when a particular variable must always precede an effect. This effect need not be the sole result of the one variable. A cause is termed "sufficient" when a particular variable inevitably initiates or produces an effect. Any given cause may be necessary, sufficient, neither, or both.	6
Clinical pathway or care pathway	A methodology for the mutual decision making and organization of care for a well-defined group of patients during a well-defined period.	10
Cluster randomized controlled trial	Randomized controlled trial that groups patients according to a variable such as clinic site or community and then randomizes them as a group to the intervention or comparison. This design is useful when evaluating health services or when randomization on the individual level is not possible.	8
Cohort Study	Nonrandomized study of a group of patients with a common condition or exposure over time. Data may be collected and evaluated prospectively or retrospectively.	8
Co-learning	A form of learning with and in communities, grounded in the direct attempt to challenge power relations between dominant and oppressed groups and the notions of expert and novice, teacher and learner.	11
Common data model	A common data model (CDM) specifies necessary data items that need to be collected and shared across participating institutes and clearly represents these associations and relationships among data elements and promotes correct interpretation of the data content.	

Terms	A number of words or phrases are used throughout this Report. Here are brief definitions for them.	Reference
Comparative effectiveness	A comparison of the effectiveness, safety or outcomes of different options of diagnosing or treating a given medical condition for a particular set of patients.	6
Comparator	An investigational marker (i.e., active control), or placebo, used as a reference in a study or clinical trial	12
Confidence interval	The limits within which the parameters of a study are expected to lie, provided by the standard error rate of the study. For example, a 95 percent confidence interval describes the range of values that has a 95 percent probability of containing the true value. When describing the likelihood that and intervention or exposure had an actual effect, a confidence interval that contains 1 indicates that in fact no effect may have been observed. See p-value.	13
Confounding	Confounding occurs when there is a relationship between an exposure, the outcome of interest, and a third factor called a "confounder" or a "confounding variable". Due to the presence of the confounder, it is not possible to accurately assess the relationship between the exposure and outcome of interest. A confounder must: 1) be related to the outcome and the exposure; 2) have a different distribution between exposure and non-exposure. Confounding may be accounted for by randomizing study participants, matching participants by likely confounders (such as age and sex), or controlling the effect of the confounder by stratifying the analysis.	6
CONSORT	Consolidated Standards of Reporting Trials. The main product of CONSORT is the CONSORT Statement, which is an evidence-based, minimum set of recommendations for reporting randomized controlled trials.	14
Covariate	A variable that may predict an outcome.	15
Crossover Trial	A trial in which patients act as their own controls and receive a sequence of treatments over periods of time crossing over to an alternative therapy as part of the sequence.	8
Data dictionary	A descriptive list of names (also called representations or displays), definitions, and attributes of data elements to be collected in an information system or database. The purpose of the data dictionary is to standardize definitions and ensure consistency of use.	16
Data integrity	The assurance that data are complete, verified and unaltered.	17
Decision aid	A tool that endeavors to present patients with the benefits and harms of alternative courses of action in a manner that is quantitative, comprehensive, and understandable.	4
De-identification	A process whereby information that could identify a patient is removed from a record.	6
Delayed start trial	Randomized controlled trial conducted in two phases to evaluate whether an intervention acts by reducing symptoms (Phase I) or by modifying disease (Phase II). In Phase I, patients receive either the intervention or control; in Phase II, all patients receive the intervention.	8

Terms	A number of words or phrases are used throughout this Report. Here are brief definitions for them.	Reference
Dissemination	To spread widely. Dissemination of research refers to communicating the results to specific audiences.	
Empirical data	Information based on direct evidence, experiences, or observations rather than on reasoning, theory, or arguments.	6
Engagement	Active involvement in an activity. In research, may include outreach activities between communities of researchers and communities of patients, and/or involvement of patients in the research process.	18
Evidence hierarchy	A formal system of organizing strength of evidence to support information or conclusions, from a low level (expert opinion) to a high level (randomized clinical trials, systematic reviews of randomized clinical trials). Empirical evidence generated through randomized controlled clinical trials are assumed to be the one of the highest levels of evidence, with systematic reviews and meta-analyses across multiple randomized clinical trials sometimes considered an even higher level (Guyatt 2000).	19
Experimental study	A study in which the investigator manipulates or controls a variable, like an intervention, and observes the effect on an outcome.	20
Explanatory clinical trial	Randomized controlled trial designed to determine whether a clinical intervention is effective under optimal circumstances. This is achieved by using rigorous inclusion and exclusion criteria and study protocols.	8
Exposure	In studies of health, exposure is the amount or duration that a study participant or population is in the presence of or affected by a variable being studied. The variable is typically a treatment or environmental condition that is expected to have an effect on the study participants.	21
Futility	A determination that a treatment or other intervention being studied is not producing a benefit.	22
Gap analysis	A process for defining the difference between an existing condition and a desired state. For example, a gap analysis may describe the difference between common clinical practice and the ideal practice. Often used to describe evaluation of a specific topic or evidence base to identify where evidence is weak or lacking.	23
Hard-to-reach	A general term for individuals or communities who are less likely to be involved in research because of differences or barriers that impede communication or collaboration with researchers. These barriers include language, education, social class, ethnicity, race, culture, geography, physical or cognitive impairments and other differences. A lack of trust (including prejudice and fear of potential legal or social consequences of engaging with researchers) may add to these barriers.	24
Harmonize	To transform data from different sources in a way that allows them to be treated as equivalent.	

Terms	A number of words or phrases are used throughout this Report. Here are brief definitions for them.	Reference
Heterogeneity of treatment effect (HTE)	The fact that when given the same treatment different people may respond differently.	25
Impute; imputation	The process of substituting an estimated answer into a field or data record that has missing data, or has an implausible or incorrect value.	6
Informant	There are different uses of this term in research. For this report, an informant is a patient representative who is in some way of the specific patient population that is the focus for the research project. Distinct from a participant who is a source of study data.	6
Instrumental variable	An instrument variable must have two properties: (1) It must be correlated with the suspected endogenous explanatory variable, preferably highly so. (2) It must not be correlated with the error term, e.g., it must not affect the dependent variable in any way except through the endogenous explanatory variable. If a variable is a valid instrument, the coefficients on the explanatory variables obtained from instrumental variable estimation will be unbiased.	6
Interaction test	A test of whether the interaction between the treatment indicator and the subgroup variable is statistically significant	26
Interim analysis,	Any examination of the data obtained in a study while that study is still ongoing.	2
Interpretability	Whether a score or measure indicates a meaningful difference between patients or a meaningful change over time within individuals. Generally established through reference to other measures.	27
Intervention	A treatment intended to change a variable.	
Investigator	Researcher	
Meta-analysis	A quantitative method for combining the results of multiple studies (on the same topic) to obtain an overall estimate of a particular treatment or intervention.	6
Metadata	Data about other data. For example, the number of patients included in the electronic records of a clinic or research team. Metadata permit analysis and interpretation about the study process and may permit combining information across different studies.	28
negative test result	A negative test result indicates the absence of the thing being tested for. For example, a negative HIV test indicates the absence of detectable infection with HIV. A negative result may also refer to a study that fails to demonstrate an expected result.	
N-of-1 randomized controlled trial	A crossover trial that compares two or more treatment options for a single patient. The order of therapy is assigned randomly, and ideally both the patient and clinician are blinded.	8

Terms	A number of words or phrases are used throughout this Report. Here are brief definitions for them.	Reference
Nomenclature	A naming system or list of names. Establishing a nomenclature reduces the likelihood of confusion or miscommunication. Application of naming conventions. Comparable to vocabulary, terminology	12
Nominal coverage properties	Where error of the information reported on the property can be quoted as a percentage of disagreement of valid type I error rates.	29
Non-inferiority study	A study intended to determine whether one intervention is at least as effective as another. This type of study is not designed to test whether one intervention is superior to another	
Non-randomized design	A study that does not use randomization to assign participants to intervention arms.	
Observational study	A type of study in which individuals are observed or certain outcomes are measured. No attempt is made to affect the outcome (for example, no treatment is given).	30
Outcome	Any change measured by a study. Health outcomes include length of life,, symptoms, function, and health-related quality of life. Surrogate outcomes, such as changes in blood test results or other lab values, are health outcomes not directly experienced by participants. Events or experiences that clinicians or investigators examining the impact of an intervention or exposure measure because they believe such events or experiences may be influenced by the intervention or exposure.	12
Parameter	A measurable, numeric characteristic about the population of interest, such has the mean value of some a variable	6
Participant	Someone who takes part in a research study, including being observed and measured. A person who is a source of study data. Distinct from an informant, who may be from the same population of interest as study participants, but who is involved in helping to shape or monitor a study.	12
Patient	Any individual with or at risk of a specific health condition.	
Patient Surrogate	An individual designated to provide the patient perspective or viewpoint, usually in cases when the patient cannot reliably and/or accurately report for him or herself.	4
patient-centered	Responsive to individual preferences, needs, and values	32
PCOR	Patient-centered outcomes research	
PCOR stakeholder	Anyone affected by health decisions. Stakeholders may include patients, clinicians, caregivers, policymakers.	
PCORI	Patient-Centered Outcomes Research Institute	

Terms	A number of words or phrases are used throughout this Report. Here are brief definitions for them.	Reference
Peer review	The evaluation by knowledgeable others about the technical and scientific quality and accuracy of research plans or products, or the evaluation by knowledgeable others of the effectiveness and efficiency of services ordered or performed by clinicians. Frequently, peer review refers to review of research by other researchers.	6
Phase I (phase 1)	Researchers test a new drug or treatment in a small group of people for the first time to evaluate its safety, determine a safe dosage range, and identify side effects. In the regulatory requirements for trials, the earliest trials of an intervention.	33
Phase II (phase 2)	The drug or treatment is given to a larger group of people to see if it is effective and to further evaluate its safety.	33
Phase III (phase 3)	The drug or treatment is given to large groups of people to confirm its effectiveness, monitor side effects, compare it to commonly used treatments, and collect information that will allow the drug or treatment to be used safely.	33
Phase IV (phase 4)	Studies are done after the drug or treatment has been marketed to gather information on the drug's effect in various populations and any side effects associated with long-term use.	33
PICOT	PICOT is an acronym for the elements of the clinical question: patient population (P), intervention or issue of interest (I), comparison intervention or issue of interest (C), outcome(s) of interest (O), and time it takes for the intervention to achieve the outcome(s) (T).	34
Positive test result	A positive test result indicates the presence of the thing being tested for. For example, a positive HIV test indicates detectable infection with the virus. A positive test result may also refer to a study that demonstrates an expected result.	
Post hoc	From the Latin for “after this”. Usually referring to determining a method for analyzing data after the data has already been collected.	
Power	The ability of a statistical test to reject the null hypothesis when it is truly false, In other words, power is the ability of the test to detect the true relationship as a function of the parameter value under the alternative hypothesis. A statistical test is considered to have "high" power if the probability of making a type II error is low.	6
Pragmatic Clinical Trial	Randomized controlled trial designed to determine the risks, benefits, and costs of an intervention as it would occur in routine clinical practice. This is achieved by including a broader range of patients, study sites, and outcomes to address needs of patients, clinicians, and other stakeholders.	8
Primary data	Data that are collected specifically for a study, as distinct from data found in existing records.	4
Prior (Informative prior, Non-informative)	The distribution of a parameter before an intervention. An informative prior predicts the outcomes in some way. A non-informative prior does not offer any information about the outcome.	6
Protocol	The specified plan for a study.	6

Terms	A number of words or phrases are used throughout this Report. Here are brief definitions for them.	Reference
Psychometric	Measurement of psychological variables. Measurement characteristics of scales that assess human psychological characteristics	12
P-value	A measure of "statistical significance" based on probability. Most often, if the P value is less than a specified level chosen prior to the study (typically set at 0.01 or 0.05), then the null hypothesis is rejected. If the null-hypothesis is rejected, the P value represents the likelihood that the observed difference between the intervention and control groups was obtained by chance alone.	6
QUADAS	**Qu**ality **A**ssessment of **D**iagnostic **A**ccuracy **S**tudies. A tool for assessing diagnostic accuracy studies.	35
Qualitative research	Qualitative research focuses on social and interpreted, rather than quantifiable, phenomena and aims to discover, interpret, and describe rather than to test and evaluate. Qualitative research makes inductive, descriptive inferences to theory concerning social experiences or settings, whereas quantitative research makes causal or correlational inferences to populations. Qualitative research is not a single method but a family of analytic approaches that rely on the description and interpretation of qualitative data. Specific methods include, for example, grounded theory, ethnography, phenomenology, case study, critical theory, and historiography.	4
Quantitative research	The investigation of phenomena that lend themselves to test well-specified hypotheses through precise measurement and quantification of predetermined variables that yield numbers suitable for statistical analysis.	4
Randomization	When referring to an experiment or clinical trial, the process by which animal or human subjects are assigned by chance to separate groups that compare different treatments or other interventions. Randomization gives each participant an equal chance of being assigned to any of the groups.	30
Randomized clinical trial	A study in which the participants are assigned by chance to separate groups that compare different treatments; neither the researchers nor the participants can choose which group. Using chance to assign people to groups means that the groups will be similar and that the treatments they receive can be compared objectively. At the time of the trial, it is not known which treatment is best. It is the patient's choice to be in a randomized trial.	30
Registry	A systematic collection of a clearly defined set of health and demographic data for patients with specific health characteristics, held in a central database for a predefined purpose.	36
Re-identification	A process whereby information that was de-identified is reconnected with the identity of a patient.	
Research	A systematic investigation, including research development, testing and evaluation, designed to develop or contribute to generalizable knowledge.	37

Terms	A number of words or phrases are used throughout this Report. Here are brief definitions for them.	Reference
Research Engagement	Specific bi-directional exchange between researchers and stakeholders	31
Researcher	Individual with experience and specialized training in research methods.	
Retrospective study	A research study based on data that are already collected at the time of initiation of the study. (NCI: retrospect: Looking back at events that have already taken place.) NCI: A study that compares two groups of people: those with the disease or condition under study (cases) and a very similar group of people who do not have the disease or condition (controls). Researchers study the medical and lifestyle histories of the people in each group to learn what factors may be associated with the disease or condition. For example, one group may have been exposed to a particular substance that the other was not. Also called case-control study.	30
Secondary data	Available data collected for other purposes that are then used in a study. Examples include census data, general clinical records or public health reports. Data that were collected prior to the current analysis and perhaps for a purpose different from the current analysis.	6
Selection bias	Sample estimates that do not generalize to estimates for the population of interest because the study sample was not obtained from the population of interest by simple random sampling.	6
Sensitivity	The proportion of people with a positive test result among those with the target condition.	4
Sensitivity analysis	Any test of the stability of the conclusions of a health care evaluation over a range of probability estimates, value judgments, and assumptions about the structure of the decisions to be made. This may involve the repeated evaluation of a decision model in which one or more of the parameters of interest are varied.	
Software validation	Confirmation by examination and provision of objective evidence that software specifications conform to user needs and intended uses, and that the particular requirements implemented through software can be consistently fulfilled.	38
Software version control	The management of source code, documents, graphics and related files in a large software project. Version-control software provides a database that is used to keep track of the revisions made to a program by all the programmers and developers involved in it.	39
Specificity	The proportion of people who are truly free of a designated disorder who are so identified by the test. The test may consist of, or include, clinical observations.	4
Stakeholder	Anyone affected by health decisions. Stakeholders may include patients, clinicians, caregivers, policymakers.	

Terms	A number of words or phrases are used throughout this Report. Here are brief definitions for them.	Reference
Standard error	The standard deviation (a measure of the variability or dispersion of data.) of an estimate of a population parameter. The standard error of the mean is the standard deviation of the estimate of the population mean value.	4, 6
STARD	**STA**ndards for the **R**eporting of **D**iagnostic accuracy studies. The objective of the STARD initiative is to improve the accuracy and completeness of reporting of studies of diagnostic accuracy, to allow readers to assess the potential for bias in the study (internal validity) and to evaluate its generalizability (external validity). The STARD statement consists of a checklist of 25 items and recommends the use of a flow diagram which describe the design of the study and the flow of patients.	40
Statistical analysis plan	A comprehensive and detailed description of the methods for, and presentation of, data analyses for a study protocol. The plan ensures that analyses are conducted in a scientifically valid manner and that decisions are documented.	41
Statistical significance	Describes a mathematical measure of difference between groups. The difference is said to be statistically significant if it is greater than what might be expected to happen by chance alone.	30
Study	An investigation of a research question. A study may be observational or experimental.	12
Superiority trial	A trial with the primary objective of showing that the response to the investigational product is superior to a comparative agent (active or placebo control).	42
Transformation logic	The formulas or rules (as codified in programming code) that govern the way data are mapped across multiple data sets.	43, 44
Trial	A prospective biomedical or behavioral research study of human subjects that is designed to answer specific questions about biomedical or behavioral interventions (drugs, treatments, devices, or new ways of using known drugs, treatments, or devices).	45
Type I error	Rejection of the null hypothesis when it is actually true. Also known as a false positive or alpha error.	6
Type II error	Failure to reject the null hypothesis when it is in fact false. Also known as a false negative or beta error.	6
Unblinded analyses	Unblinded analyses are those in which the treatment group assignments of subjects are known and used in some manner in the analysis, usually (but not always) as a formal comparison between treatment groups.	2
Utility	Patient preferences that are measured with techniques consistent with modern utility theory. Patient preferences refer to the degrees of subjective satisfaction, distress, or desirability that patients or potential patients associate with a particular health outcome. Utility theory is based on specific axioms that describe how a rational decision maker ought to make a decision when the outcomes of that decision are uncertain. Commonly used measures of utility include the “standard gamble” or “time trade-off” techniques.	4

Appendix F References

1. http://www.merriam-webster.com

2. http://www.fda.gov/downloads/DrugsGuidanceComplianceRegulatoryInformation/Guidances/UCM201790.pdf

3.http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf

4 http://jamaevidence.com/glossary

5. Adapted from CDISC Clinical Research glossary:
http://www.cdisc.org/stuff/contentmgr/files/0/be650811feb46f381f0af41ca40ade2e/misc/cdisc_2009_glossary.pdf

6. http://www.hsrmethods.org/glossary.aspx

7.http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/ucm071121.pdf

8. Study Designs for Assessing the Effectiveness of Therapeutic Interventions

9. http://www.fda.gov/downloads/Drugs/GuidanceCompliance

10. European Pathway Association: www.e-p-a.org

11. New Directions for Adult and Continuing Education V 2000, Issue 87.
onlinelibrary.wiley.com/doi/10.1002/ace.8708/pdf

12. CDISC Clinical Research glossary
http://www.cdisc.org/stuff/contentmgr/files/0/be650811feb46f381f0af41ca40ade2e/misc/cdisc_2009_glossary.pdf

13. Adapted from Combined, Research Methodology: Methods and Techniques by C. R. Kothari
http://books.google.com

14. http://www.consort-statement.org/

15. http://covariate.askdefine.com/

16. AHIMA (American Health Information Management Association)

17. http://cflegacy.research.umn.edu/datamgtq1/MDI_020.html

18. OHSU (meaning the contractor report that is also cited in the main text)

19. Adapted from Observational versus Experimental Studies: What's the Evidence for a Hierarchy
http://www.ncbi.nlm.nih.gov

20. Adapted from Hulley SB, Feigal D, Martin M, Cummings SR. Designing a New Study: IV. Experiments. Chapter 11 in Designing Clinical Research. An Epidemiologic Approach. Hulley SB and Cummings SR (Eds). Baltimore: Williams and Wilkins. 1988.

21. Adapted from Journal of Exposure Analysis and Environmental Epidemiology
http://www.nature.com/jes/journal/v15/n1/pdf/7500411a.pdf

22. Adapted from http://www.ncbi.nlm.nih.gov/pubmed/16159066

23. University of Toronto Faculty of Medicine
http://www.cepd.utoronto.ca/newsletter/2009_fall/Quick_Tips_initiating_a_learning_needs_assesment__Gap_Analysis_of_health_care_issues.pdf

24. Univ of Md, OHSU (contractor reports)

25. Adapted from the text.

26. From the standard.

27.

http://www.esourceresearch.org/eSourceBook/PatientReportedOutcomes/8Interpretability/tabid/178/Default.aspx

28. Adapted from http://www.techterms.com/definition/metadata

29. Adapted from The Bootstrap and Edgeworth Expansion by Peter Hall. Books.google.com

30. http://www.cancer.gov/dictionary

31. OHSU (contractor reports)

32. IOM (re: care but applicable to research).

33. http://www.nlm.nih.gov/services/ctphases.html

34.http://journals.lww.com/ajnonline/Fulltext/2010/03000/Evidence_Based_Practice,_Step_by_Step__Asking_the.28.aspx

35. http://www.biomedcentral.com/1471-2288/3/25

36. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC349377/

37. US Dept of Health and Human Services

http://www.nichd.nih.gov/health/clinicalresearch/references/upload/clinical_research-definitions_procedures.pdf

38. FDA http://www.fda.gov/RegulatoryInformation/Guidances/ucm126954.htm

39. PC Magazine http://www.pcmag.com/encyclopedia_term/0,1233,t=version+control&i=53750,00.asp

40. http://www.stard-statement.org/

41. NIH via http://www.sctweb.org/presentations/2007/SCT-SAP-HANDOUT-01May2007pdf.pdf

42. FDA http://www.fda.gov/RegulatoryInformation/Guidances/ucm127509.htm

43. Stanford.edu http://www.stanford.edu/dept/itss/docs/oracle/10g/server.101/b10736/transform.htm

44. Oracle http://docs.oracle.com/cd/B28359_01/server.111/b28313/transform.htm

45. NIH http://grants.nih.gov/grants/glossary.htm#C

Appendix G: Acknowledgments

Principal Investigators and Research Team Members

University of Maryland, Pharmaceutical Health Services Research Department (Daniel Mullins, Ph.D.)
Beth Barnet, MD
Susan DosReis, PhD
Liz Jansky (Westat)
Karen Kauffman, Ph.D.
Ebere Onukwugha, Ph.D.

Mayo Clinic, Knowledge and Evaluation Research Unit (M. Hassan Murad, M.D., MPH)
Juan Pablo Domecq Garces, M.D.
Gabriela J. Prutsky Lopez, M.D.
Zhen Wang, Ph.D.
Tarig A. Elraiyah, MBBS
Mohammed Nabhan, M.D.
Juan P. Brito Campana, MBBS
Kasey Boehmer, B.A.
Rim Hasan, M.D.
Belal Firwana, M.D.
Nathan Shippee, Ph.D.
Jeff A. Sloan, Ph.D.
David T. Eton, Ph.D.
Patricia J. Erwin, MLS
Victor M. Montori, M.D., MSc

Oregon Health & Science University, The Center for Evidence-Based Policy (Pam Curtis, MS)
Beth Church
Cathy Gordon, MPH
Valerie King, MD, MPH
Catherine Pettinari, PhD
Katharine Ryan, MPH
Samantha Slaughter-Mason, MPH
Aasta Thielke, MPH

Oxford Outcomes, Ltd., Patient Reported Outcomes (Andrew Lloyd, PhD)
Sarah Acaster, MSc
Tricia Cimms, BS

Northwestern University/UNC Chapel Hill (Zeeshan Butt, PhD /Bryce Reeve, PhD)

Johns Hopkins University (Tianjing Li, MD, MHS, PhD)
Susan Hutfless, SM, PhD
Kay Dickersin, MA, PhD
Daniel Scharfstein, ScD
James Neaton, PhD
Joseph Hogan, ScD
Roderick Little, PhD
Michael Daniels, ScD
Jason Roy, PhD
Vincent Mor, PhD, MED
Andrew Law, MHS

Johns Hopkins University – School of Medicine (Ravi Varadhan, PhD)
Elizabeth A. Stuart, PhD
Thomas A. Louis, PhD
Jodi B. Segal, MD, MPH
Carlos O. Weiss, MD, MHS

Berry Consultants (Scott M. Berry, PhD)
Michelle A. Detry, PhD
Roger J. Lewis, MD, PhD
Kristine R. Broglio, MS
Jason T. Connor, PhD
Donald A. Berry, PhD

Brown University (Constantine Gatsonis, PhD)
Ruth Carlos, MD
Ilana Gareen, PhD
Jeremy Gorelick, PhD
Larry Kessler, ScD
Joseph Lau, MD
Carolyn Rutter, PhD
Christopher Schmid, PhD
Anna N. A. Tosteson, ScD
Thomas Trikalinos, MD, PhD

Brigham and Women's hospital and Harvard Medical School (Josh Gagne, PharmD, ScD)
Jennifer M Polinski, ScD, MPH
Jerry Avorn, MD
Robert J Glynn, PhD, ScD
John D Seeger, PharmD, DrPH

Outcome Sciences, Inc. (A Quintiles Company) (Richard Giklich, MD)
Nancy A. Dreyer, PhD, MPH
Michelle B. Leavy, MPH
Priscilla Velentgas, PhD
Laura Khurana

University of California San Diego (UCSD) (Lucila Ohno-Machado, MD, PhD)
Michele E. Day, PhD
Robert El-Kareh, MD, MEng, MPH
Xiaoqian Jiang
Hyeon-Eui Kim, RN, PhD
Daniella Meeker
Dennis Browe
Katherine Kim

Hayes, Inc. (Petra Nass, PhD)
Susan Levine, DVM, PhD

NORC at the University of Chicago (David Rein, PhD)
Wilhelmine Miller, PhD (contributor)

Duke Evidence-Based Practice Center (Evan Myers, MD, MPH)
Amanda J. McBroom, PhD
Lan Shen, MD
Rachael E. Posey, MSLS
Rebecca Gray, DPhil
Gillian D. Sanders, PhD

Medical College of Wisconsin (Theodore Kotchen, MD)
Ryan Spellecy, PhD

Cecil G Sheps Center for Health Services Research
University of North Carolina at Chapel Hill; Chapel Hill, NC (Tim Carey, MD, MPH)
Amica Yon, PharmD
Chris Beadles, MD
Roberta Wines, MPH