Methods Future Research Needs Report

Number 8

Framework for Considering Study Designs for Future Research Needs

Agency for Healthcare Research and Quality
Advancing Excellence in Health Care • **www.ahrq.gov**

Methods Future Research Needs Report

Number 8

Framework for Considering Study Designs for Future Research Needs

Prepared for:
Agency for Healthcare Research and Quality
U.S. Department of Health and Human Services
540 Gaither Road
Rockville, MD 20850
www.ahrq.gov

Contract No. 290-2007-10056-I

Prepared by:
RTI-UNC Evidence-based Practice Center
Research Triangle Park, NC

Investigators:
Timothy S. Carey, M.D., M.P.H.
Gillian D. Sanders, Ph.D.
Meera Viswanathan, Ph.D.
Thomas A. Trikalinos, M.D., Ph.D.
Elisabeth Kato, M.D., M.R.P.
Stephanie Chang, M.D., M.P.H.

AHRQ Publication No. 12-EHC048-EF
March 2012

This report is based on work conducted by the Research Triangle Institute International-University of North Carolina Chapel Hill (RTI-UNC) Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2007-10056-I). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care researchers and funders of research make well informed decisions in designing and funding research and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of scientific judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical research and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances.

This report may be used, in whole or in part, as the basis for research design or funding opportunity announcements. AHRQ or the U.S. Department of Health and Human Services endorsement of such derivative products or actions may not be stated or implied.

Persons using assistive technology may not be able to fully access information in this report. For assistance contact EffectiveHealthCare@ahrq.hhs.gov.

None of the investigators have any affiliations or financial involvement that conflicts with the material presented in this report.

Suggested citation:
Carey T, Sanders GD, Viswanathan M, Trikalinos TA, 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. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

The Agency for Healthcare Research and Quality (AHRQ), through its Evidence-based Practice Centers (EPCs), sponsors the development of evidence reports and technology assessments to assist public- and private-sector organizations in their efforts to improve the quality of health care in the United States. The reports and assessments provide organizations with comprehensive, science-based information on common, costly medical conditions, and new health care technologies and strategies. The EPCs systematically review the relevant scientific literature on topics assigned to them by AHRQ and conduct additional analyses when appropriate prior to developing their reports and assessments.

An important part of evidence reports is to not only synthesize the evidence, but also to identify the gaps in evidence that limited the ability to answer the systematic review questions. AHRQ supports EPCs to work with various stakeholders to identify and prioritize the future research that decisionmakers need. This information is provided for researchers and funders of research in these Future Research Needs papers. These papers are made available for public comment and use and may be revised.

AHRQ expects that the EPC evidence reports and technology assessments will inform individual health plans, providers, and purchasers as well as the health care system as a whole by providing important information to help improve health care quality. The evidence reports undergo public comment prior to their release as a final report.

We welcome comments on this Future Research Needs document. They may be sent by mail to the Task Order Officer named below at: Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, or by email to epc@ahrq.hhs.gov.

Carolyn M. Clancy, M.D.
Director
Agency for Healthcare Research and Quality

Jean Slutsky, P.A., M.S.P.H.
Director, Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Stephanie Chang, M.D., M.P.H.
Director, EPC Program
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Elisabeth Kato, M.D.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Acknowledgments

The authors gratefully acknowledge the following individuals for their assistance with this manuscript: Roberta Wines, M.P.H., Research Associate at the University of North Carolina Chapel Hill Cecil G. Sheps Center for Health Services Research; Sharon Barrell, editor, RTI International; and Loraine Monroe, publications specialist, RTI International.

Framework for Considering Study Designs for Future Research Needs

Abstract

The Agency for Healthcare Research and Quality (AHRQ) commissioned a series of methods papers to describe and provisionally recommend methods for the Future Research Needs (FRN) projects and reports. Our paper proposes a method for composing the study design considerations portion of FRN reports; this portion discusses proposed research designs for addressing the high-priority research needs as determined by stakeholder opinion. Each proposed research design is accompanied by a discussion of advantages and disadvantages for using the approach. Here we present a framework based on a standard taxonomy for study designs and criteria for evaluating the appropriateness of a study design to address a particular research need. Criteria for evaluating each design include the advantages of the study design for producing a valid result; the resources needed and duration of the proposed study; the availability of data or ability to recruit; and ethical, legal, or social issues. The study design discussions are intended to be considerations and not prescriptive to researchers or funders. This framework is intended to help FRN project teams consistently apply criteria to determine which study design may be most appropriate for each research question, and to provide suggestions for the rationale and presentation of their design choices.

Contents

Tables

Figures

Appendixes

Introduction

This methods paper was commissioned by the Agency for Healthcare Research and Quality (AHRQ) as one of a series of papers addressing methods issues in the relatively new area of explicit discussion of future research needs (FRN) as part of comparative effectiveness research (CER). This paper is intended to reflect current and recommended practices for the AHRQ Evidence-based Practice Centers (EPCs); it represents initial guidance, as additional experience with future research needs may lead to modifications in the future. Other papers in this methods series on FRN in CER may be found on AHRQ's Effective Health Care (EHC) Program Web site: www.effectivehealthcare.ahrq.gov/futureresearchneedsmethods.cfm.

Comparative Effectiveness Research

CER is the "generation and synthesis of evidence that compares the benefits and harms of alternative methods to prevent, diagnose, treat, and monitor a clinical condition or to improve delivery of care. The purpose of CER is to assist consumers, clinicians, purchasers, and policymakers to make informed decisions that will improve care both at the individual and the population levels."[1] CER comprises a broad range of activities and types of study, encompassing systematic reviews, secondary data analyses, randomized controlled trials, prospective observational studies, health systems research, and dissemination of results to the public, providers, policymakers, and other key stakeholders. Key components of CER include comparisons between active treatments, policies, or diagnostic strategies and evidence from research conducted in settings similar to those in which most patients with a given condition are treated. The explicit nature of CER is demonstrated by the descriptions of proposed study questions through the PICOTS format, in which each key question is described using six dimensions: population, intervention, comparator treatment or test, outcomes assessed, time frame, and study setting.[2]

Future Research Needs

Systematic reviews of focused clinical and policy questions reach conclusions whenever feasible and describe the strength of evidence (SOE) supporting those conclusions. However, many reviews find only low or insufficient SOE to address a given key question. Problems are often identified with the amount or quality of the literature examined, leading to an inability to address all of the components of the key study questions to sufficiently address the clinical and policy needs that led to the key questions. Gaps in the evidence remain. A common criticism of systematic reviews is that, although they generally contain a section describing the limitations of the research just reviewed, these limitations sections often are very general (e.g., "larger trials are needed") and provide relatively little guidance to funders or the research community regarding the next study or series of studies needed to advance a given field.[3] Yet a key, and to date underused, role of the systematic review process is to stimulate new research to address identified gaps in the literature.

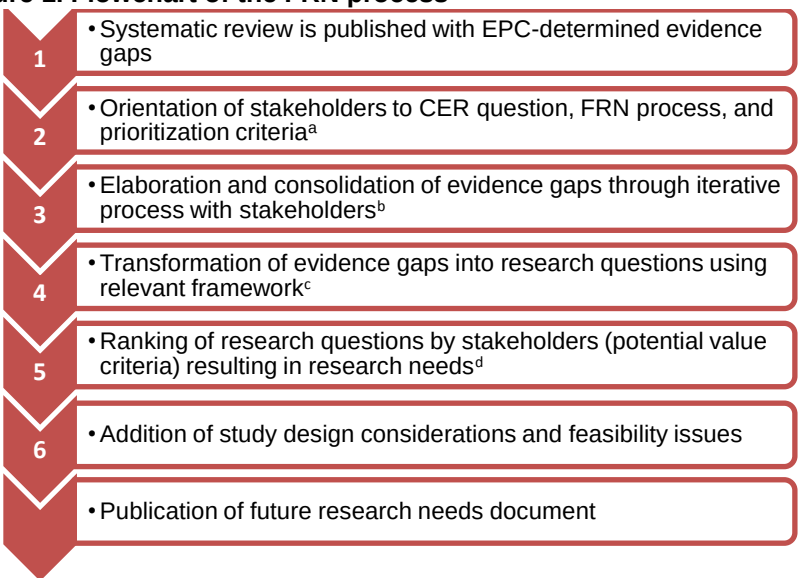
With these FRN papers and accompanying methods papers, the AHRQ EPC Program distinguishes between the evidence gaps that are identified from within a systematic review and those that are prioritized and clearly defined as research needs by stakeholders based on their potential impact on practice or care. A more explicit and prioritized listing of research needs, with guidance regarding how to address those needs, could allow the impact of systematic

reviews to be more fully realized and increase the pace of research to provide meaningful answers. The audience for FRN reports includes the research community, funders, policymakers, and advocacy groups. Reducing the time between synthesis of evidence, identification of FRNs, and initiation of studies to address those needs is urgently required in the current health care environment.

The Future Research Needs Process

In the AHRQ EHC program, FRN documents are derived from systematic reviews of CER questions. The FRN document follows online publication of the systematic review and serves as a standalone document. Figure 1 shows the flow of an FRN project.

Figure 1. Flowchart of the FRN process



[a] May include identification of additional evidence gaps.
[b] Reduction through topic consolidation, preliminary prioritization, and consideration of ongoing research (duplication criteria).
[c] Evidence gaps that address specific methods issues would not use the PICOTS framework.
[d] May require iterative steps.

Each FRN report begins by identifying a list of evidence gaps from the systematic review (in draft or final form), which may be augmented with input from a multidisciplinary panel of stakeholders familiar with both the research methods as well as the clinical and policy content of the systematic review. The EPC then works with the stakeholder group to elaborate and consolidate the evidence gaps, taking into consideration any ongoing or planned research that may already be addressing gaps. Potential research questions are then elaborated following the PICOTS framework with the exception of methodological questions, which may be organized differently.[4] Once the questions have been formalized, they are given a final prioritization by the stakeholders according to potential value criteria. The final list of 4 to 12 high-priority FRNs with specific questions including PICOTS definition (as appropriate) and potential study designs is published in a final document intended for use by researchers and funders of research (Figure 1, Step 7).

Scope of This Paper

This paper is one of a series of papers that provides recommendations and best practices on the steps in identifying and prioritizing FRNs. Other papers that address other steps in the FRN process will be posted as they are completed at www.effectivehealthcare.ahrq.gov/futureresearchneedsmethods.cfm.

Our paper proposes a method for composing the study design considerations portion of FRN reports. This portion discusses proposed research designs for addressing the high-priority research needs as determined using stakeholder input. Each proposed research design is accompanied by a discussion of advantages and disadvantages for using the approach. Here we present:

- a framework based on a standard taxonomy for study designs and criteria for evaluating the appropriateness of a study design to address a particular research need and
- criteria for evaluating each design: the advantages of the study design for producing a valid result; the resources needed and duration of the proposed study; the availability of data or ability to recruit; and ethical, legal, or social issues.

The study design discussions are intended to be considerations and not prescriptive to researchers or funders. This framework is intended to help FRN project teams consistently apply criteria to determine which study design may be most appropriate for each research question and provide suggestions for their rationale and presentation.

Role of Study Design Considerations Within Future Research Needs

The EPCs add a short discussion of study design considerations specific to each of the highest-priority research needs after stakeholders prioritize the needs (see Step 6 in Figure 1). The purpose of this discussion in the FRN document is to leverage the knowledge gained in the systematic review process and assist funders, researchers, and stakeholders in further describing how a subsequent research project, potentially initiated by funders, advocates, or policymakers and conducted by researchers, might approach meeting an FRN. Adding the study design considerations to the FRN document takes advantage of the time and expertise dedicated to future research by the EPC team and stakeholders in developing the needs. Because more than one study design may be appropriate to address a given research need, these are simply considerations and are not intended to be prescriptive. The discussion of study design in the AHRQ FRN documents should be only a first step in developing a research project, giving funders, policymakers, and researchers some initial ideas that will hopefully stimulate discussion and additional efforts.

To maintain consistency across the future research papers and to communicate clearly with stakeholders regarding FRN projects, terms must be clearly defined and a set of basic dimensions must be considered across all reports, although the conclusions will vary according to the specific topic. Therefore, this paper proposes terminology and methods of presentation of FRN study design consideration materials to provide consistency and a common language for future reports. This document presents the early stage of development for these methods; we expect that they will be refined in the months and years to come.

From Systematic Review to Study Design Considerations

As noted above, moving from systematic review to study design considerations starts with the identification of evidence gaps and their prioritization into research needs. At this point, the research needs can be further analyzed and characterized to become "operationalized." A framework for doing so is presented in "Frameworks for Determining Research Gaps During Systematic Reviews."[5] The process of characterizing gaps may reveal important considerations for study design. The framework has two steps.

The first step is to assess why existing evidence is inadequate, using criteria based on the same criteria that are used for assessing SOE.[6]

- Insufficient or imprecise information (corresponding to simple lack of studies or the GRADE* concept of precision)
- Biased information (corresponding to bias)
- Inconsistent or unknown consistency results (corresponding to consistency)
- Not the right information (corresponding to directness or applicability)

The second step is to apply PICOTS criteria to the identified evidence gaps, specifying the population of interest, the proposed intervention, the comparator treatment test or policy, the outcomes to be assessed, the timeframe for the study, and the setting of the research. Most research needs can be fully characterized with this framework, although research needs that are primarily methods enhancements (such as development of outcomes measures or a statistical technique) may not be able to be placed in such a schema. Once the research needs have been fully characterized, it is possible to make suggestions about study designs. For example, if the problem is a simple lack of data, then a wide range of trial designs may be able to add to the picture. However if the problem is lack of precision, power may be an important factor in determining what kind of study will be able to answer the question. If bias is the problem, studies that replicate the same methodological problems are unlikely to resolve the question. This may be a problem of poorly designed or executed studies, or it may have to do with the type of study design. For example, existing studies that are mainly observational for a topic may generate concern regarding unmeasured confounding. Inconsistency may be due to unidentified heterogeneity in the population or intervention, or it could be due to lack of consensus on what the study measures should be. In the first case, it will be important to identify possible causes of heterogeneity and incorporate them into the new study design (for example, using inclusion criteria to create homogeneity or through stratification). If the problem is "not the right information," additional considerations come to mind. If a question cannot be answered because all existing studies only measured surrogate markers, additional studies on surrogate markers are unlikely to move the field forward. Or if studies have only been carried out in academic settings but the intervention will be implemented in primary care, then a focus on use of primary care networks would be useful. Study designs other than trials may be appropriate when methods

---

*The GRADE (Grading of Recommendations Assessment, Development and Evaluation) working group has developed an approach to grading quality of evidence and strength of recommendations in health care. Its Web site is at www.gradeworkinggroup.org.

enhancements are needed or very long-term outcomes need to be assessed in highly generalizable populations.

Throughout this and other FRN documents, we propose common terminology across EPCs and topic areas to aid in communication across disciplines.

Common Terminology for Study Designs

An initial task is to categorize study designs, so that the EPCs will be consistent in their terminology across reports and to use common descriptors of the potential advantages and disadvantages of each study design when applied to a given FRN. The following categories of studies and their descriptions are derived from the AHRQ report "Developing and Testing a Tool for the Classification of Study Designs in Systematic Reviews of Interventions and Exposures" published in December 2010.[7] We have added several additional considerations of study types that might be useful in study design consideration discussions. Although these descriptions are consistent with other study design categorization systems, the authors of the report found that the agreement across raters (kappa statistics) in classifying studies was fair at best.[7] Researchers from different disciplines (epidemiology, health policy, etc.) may use differing terms to describe the same study design, and a more consistent terminology will assist communication across disciplines and reports. Different disciplines may use differing terminology to describe the similar study designs. The list below is not intended to be comprehensive and modifications in design and terminology will be appropriate in some circumstances. The terms below do, however, represent a common point of communication prior to adaptation to each study question. The study design terms used in this discussion are listed in Table 1, and descriptions are provided in Appendix A.

Table 1. Study design terms

Terms	
A. Randomized controlled trial (RCT)	H. Nested case-control study
B. Nonrandomized comparative trial	I. Case-control study
C. Prospective cohort study	J. Interrupted time series (without a comparison group)
D. Retrospective cohort study	K. Before-after study
E. Interrupted time series with comparison group	L. Cross-sectional study
F. Controlled before-after study	M. Noncomparative study
G. Nonconcurrent cohort study	N. Other considerations: Meta-analysis of individual participant data Modeling Additional systematic review

Criteria for Considering Study Designs

To provide additional guidance for using this categorization scheme for the consideration of study designs for FRN projects, we have added brief summaries of the basic dimensions that should be considered when discussing different possible study designs. It is hoped that the process of reviewing and synthesizing the studies to date for a particular topic will give the reviewers some insight into how these criteria might affect the feasibility or utility of additional studies. The first criterion is the ability of the design to produce a valid result; the latter three criteria are elements of determining the feasibility of conducting a study. In recommending one or two study designs, FRN teams will need to consider each of these issues, balancing them against one another. Although a given study design may be optimal for producing a valid result,

feasibility issues are critical in choosing a study design. An optimal study design that, for example, fails to recruit adequate numbers of subjects or has a very high dropout rate will not add to our knowledge of a field. These descriptors and comments are generic and may need to be modified or added to for each given FRN exercise, given the clinical and policy context of the condition under consideration. In developing these summaries, we used the following working definitions.

Advantages of Study Design for Producing a Valid Result

This dimension allows for exploration of how study design contributes to validity and how the importance of different types of validity may vary according to the question and the body of previous evidence. The main advantage of the study design in providing a valid result (the greatest validity in answering a given question) is avoiding systematic error. In CER, additional advantages may include the ability to rapidly conduct a study in a policy-relevant environment and address issues of the effectiveness, as opposed to efficacy, of an intervention or diagnostic test.

Given that the context is effectiveness research, generalizability or external validity should be included in these study design considerations. Although issues of internal validity are often specific to the study design, generalizability may depend on the clinical condition under study, the settings in which it is treated, and the populations affected.

Resource Use, Size, and Duration

This dimension describes the magnitude of resources required to achieve a meaningful answer to the research question using this study design. Although this dimension will often translate into higher versus lower cost, resource use could also include the opportunity costs of practice resources or investigator time. Within study designs, costs are not fixed. Randomized controlled trials (RCTs) can be very expensive, but large, simple trials have also been conducted for relatively modest costs; trials requiring only a short duration (some infectious disease trials) are less expensive than chronic disease trials, which may need to observe subjects for years. The size of a study depends on a variety of factors, including the expected number of outcome events occurring or, in the case of a nondichotomous outcome event, the precision with which a scale can be assessed. The duration of a study depends on the anticipated accrual of recruited subjects, the incidence of outcome events, and the natural history of the condition studied.

Availability of Data and Ability To Recruit

This dimension assesses the likelihood that the study design will be able to achieve the basic inputs (data or patients) required for launch. Data and recruitment are partially dependent on study design and partially dependent on the content area. For example, subjects are generally much easier to recruit into observational studies than intervention trials. FRN researchers may know of already existing secondary databases that might address an identified evidence gap.

Ethical, Legal, and Social Issues

This dimension deals with the external issues that could affect the feasibility or desirability of different study designs. Potential problems with ethical conduct of research are likely to vary

depending on the content of the proposed measurement and intervention and the current standard of care, and may not be a consistent feature of a particular study design.

These guidelines should be applied with flexibility to considerations of study designs for FRNs. The advantages and disadvantages of a study design may change depending on the study question or the setting. The resources required for a study design depend on the intervention proposed. The features of a specific FRN may make some study designs better suited under those particular circumstances. Considerations should include an understanding of the context of the research, including the condition or health policy being investigated, the existing body of evidence, and potential utility and quality of the data gained through different study designs. The urgency with which an answer to the research question is needed may be an additional ethical consideration in prioritization of research needs.

Issues in the Application of the Framework to Future Research Needs Projects

A discussion of study design considerations may assist researchers and funders in determining whether to examine a given research question and guide the resources needed to address the research need. We anticipate that this information will provide a starting point for study planning and stimulate discussion. We do not intend that the future research considerations be prescriptive or exclude creative study designs or innovative use of existing data for CER. Some promising research options may not be considered by the EPC team and stakeholders. The FRN documents are intended to stimulate additional discussion among researchers and stakeholders, not truncate debate or planning.

Composition of the FRN Team

Addressing study designs to collect new information is a qualitatively different activity from systematic review. Systematic review requires knowledge of a range of study designs and analytic methods, as well as deep knowledge of the strengths and weaknesses of each. Part of a systematic review is acquiring knowledge of specifics of the clinical condition and its treatments. Such knowledge is intrinsic to study quality and overall strength of evidence assessments. However, the members of the FRN team may not necessarily have personal experience with conducting the study designs and especially not conducting the study design in the subject area of the report. As generalists, EPC teams bring epidemiologic methods expertise and broad perspective to the work, but experience in specific study conduct issues such as subject recruitment, prevention of subject dropout, management of cointerventions, and cross-over between study groups can be helpful. EPC core staff may not have expertise in all these areas. Recruiting an FRN team member with expertise in these study conduct areas can provide a practical balance to the general epidemiological methods expertise. Also helpful will be maintaining the focus on the key questions addressed in the original systematic review and the gaps identified in that review through the GRADE ratings and other methods. The members of the stakeholder panel can provide additional perspectives but should not be considered a substitute for such expertise on the study team. Stakeholders have subject expertise and are often experienced researchers in the area but cannot be expected to devote the time to provide detailed methods advice.

Timing of Study Design Application

Study design considerations will be developed by the EPC team after the evidence gaps have been translated into FRNs and after priorities are assigned by the FRN stakeholder group during conference calls and prioritization exercises. Stakeholders assist the EPC team in identifying potentially many evidence gaps and then prioritize the gaps into high and low priority, in recognition that not all evidence gaps can be resolved in a short period of time. Design considerations will be assigned only for the final high-priority FRNs. As the FRN projects are currently conducted, the stakeholders do not have an opportunity to review and comment on the proposed study designs until the report is posted for public comment. During the public comment period, stakeholders and other members of the public can simultaneously critique the draft study design considerations.

Weighing Different Considerations in Study Designs

Report authors will need to use judgment to arrive at FRN recommendations for study designs. We recommend that the report authors consider the factors of resource use, ethical factors, data availability and recruitment feasibility, and validity when recommending one or several designs to funders, researchers, and policymakers. The emphasis on each factor will vary depending on the health care condition or problem being examined and the specific issues raised by each FRN. We do not feel that a specific quantitative weighing or scale is appropriate to assess the overall desirability of a given study design. Such a process would be ad-hoc and justifiably subject to criticism. In addition, a substantive ethical problem would, of course, trump any validity or resource use advantage. The most prominent example of assessing study designs is the peer review process used by the National Institutes of Health and AHRQ, which assigns individualized scores to components of a fully elaborated research proposal (e.g., significance, impact, approach, investigators, environment)—the summative score is not a combination of the individual scores, but rather a standalone 1 through 9 numeric score. For FRN reports, we recommend that the authors list and assess one to three study designs for each of the final high-priority FRNs, with limited textual description—just enough that a reader can understand the rationale for the choices. The FRN team will need to strike a balance between being overly prescriptive and provocative regarding study designs and being too general and bland to be helpful to stakeholders not involved with the process.

Sample Size Calculations

One descriptor that can guide researchers and funders in planning a future research project is the anticipated number of subjects needed for the study design within a given condition. Sample size in any study depends on multiple factors: the study outcome and its frequency, the variance of the outcome when it is a scale, the anticipated dropout rate, and the difference in the study outcome between the intervention and comparator groups that is needed to be of clinical or policy importance. As part of study planning investigators do, of course, discuss sample size issues in some detail, so the use of sample size calculations in an FRN document should be considered only a first approximation. Such discussion can be useful, however, in illustrating some of the challenges inherent to CER. By definition, CER involves comparing two active treatments, tests, or policies as opposed to a no treatment or a placebo comparison group. The effect size identified between two active interventions is generally smaller than the effect size between either active intervention under consideration and a placebo or no treatment. A larger

sample size will, therefore, be needed in the CER study to detect this smaller difference. Sample size calculations in FRN documents should be regarded as illustrative, not definitive, and may not be needed for all FRNs. When sample size estimates are provided, the basis for the sample size calculations (e.g., event rates) should be referenced. An example of sample size calculations to inform feasibility of specific designs is provided at the end of this manuscript (Example 4).

Format

Either a narrative (several sentences up to one to two paragraphs) or a tabular format could be used in presenting study design considerations to readers. In the examples that follow, we present both formats. A table is most appropriate for straightforward designs, while a narrative approach is most appropriate for methods studies or for situations in which a study design choice might depend on a number of factors. For example, a community-based RCT might be an optimal study design in some circumstances, but success in such a trial might be conditional on development of brief, acceptable, and valid outcomes instruments. Such considerations would be more clearly presented in text than a table.

Discussion of Study Designs Not Included in the Systematic Review

Some study designs recommended to address FRNs may include designs that are not RCTs or controlled studies or designs that rely on retrospectively collected data. Such study designs may be useful to address specific gaps regarding generalizability of findings from RCTs and methods issues or to identify possible harms in large populations when the study samples in RCTs may be too small or not sufficiently generalizable as part of the evaluation of a drug or technology. Although projects funded under the American Recovery and Reinvestment Act of 2009 (ARRA-funded projects) focus on CER with active treatment comparators, a CER review may include findings that treatment efficacy compared with placebo has not been established or has not been established for some key subpopulations. Additional RCT efficacy trials would therefore be in order. These designs are generally within scope of the key questions of the original systematic review.

For example, several FRN projects to date have found that there is a gap in valid measurement instruments for patient-centered outcomes, and stakeholders may highly rate the need for such instruments. Measurement designs may not have been reviewed in the original review, but further CER trials may depend on the availability of adequate outcome measures. Therefore, it may be difficult to know whether this proposed gap has already been filled. The original study team may be able to determine from the background readings and title and abstract reviews whether certain study designs or research exists but was not included in the systematic review. However, some FRN teams might not have this background information if they were not authors of the original systematic review. Because substantial new systematic data collection and quality assessment of additional literature are not within the scope of the FRN process, the FRN team will not perform a detailed full text review and quality rating of additional studies. Exploring the existence of additional research with the diverse group of stakeholders may help determine if an evidence gap exists, with very limited examination of new literature. However, when it is not clear whether the proposed highly rated new research need is duplicative, this should be noted in the FRN report so that researchers considering this topic would be promoted to conduct a review of the literature to make sure that the research need has not already been

addressed. Conversely, if an identified gap could be filled by RCTs, for example, but trials have been unable to accrue subjects because of recruitment difficulties, this concern should be included in the study design considerations text. The intent of the FRN process is for these to be practical considerations, based on both research theory and implementation issues.

Examples of Study Design Considerations

The examples that follow illustrate the team's thinking about the content and format of demonstrating study design considerations within FRN reports. The first three are examples of a single FRN; the fourth example is a discussion of sample size issues. The examples are derived from the first series of EPC FRN reports; the needs have been reformatted to reflect the recommended presentation described in this paper. A tabular format is considered most succinct (Example 2; see Table 2). As discussed above, the descriptors of the study design considerations should be brief. Only one to three (rarely more) of the most feasible study designs should be presented; we believe it is redundant and potentially confusing to present iterations of why, for example, a cross-sectional study design is inappropriate as a means of filling an evidence gap related to treatment effectiveness. For some gaps, only an RCT might suffice and, therefore, only one study design presentation is appropriate. As discussed previously, the authors of the FRN report should encourage creativity and emphasize that these considerations are meant to be illustrative, not prescriptive. Further, advances in analytic methods may enable alternative study designs not anticipated by the FRN team. Additionally, the FRN project team should consider and discuss the lessons learned from studies included in the CER. A discussion of methodological weaknesses that limit the strength of available evidence could be used to support a suggestion that would prevent repeating previous mistakes. Methodological evidence gaps, if apparent, should also be addressed in the study design considerations.

Example 1: Narrative/Bulleted Text

Content area: Fixation of fractured hip: "Do certain procedures (e.g., internal fixation) work better than others for frail elder patients?"[8]

Randomized Trial

- Advantages of study design for producing a valid result: A well-done RCT will produce the most convincing results and, if inclusion/exclusion criteria and setting are realistic, should be fairly generalizable.
- Resource use, size, and duration: An RCT has to be large, because the question compares active treatments and the effect size may be small and easily swamped by other causes of morbidity and mortality in this population. Duration depends on whether the trial focuses on peri-procedural complications and short-term outcomes or on the longer-term durability of different treatment options. In either case, the resource requirements will be large or very large, given that the effect size between the treatments might be modest.
- Ethical issues: As long as equipoise exists among the treatment options, ethical issues regarding enrollment should not be present. However, if the study includes patients with dementia, consent issues may occur.

- Availability of data or ability to recruit: Recruitment may be slow, because this is a subpopulation of the population of hip fracture patients, and it may be difficult to reach large numbers.

Prospective Cohort Study

- Advantages of study design for producing a valid result: Although concern for selection bias and unmeasured confounders will always exist, the prospective design allows data for the most relevant known confounders to be collected and controlled for. Therefore, while the results will not be as definitive as an RCT, they could be informative.
- Resource use, size, and duration: This type of study still requires a large size because of potentially small treatment effects, but it would likely be less expensive than an RCT.
- Ethical, legal, and social issues: The main ethical issue is consent in the case of patients with dementia; however, because choice of treatment is not involved, it may be of less concern.
- Availability of data or ability to recruit: Recruiting patients for this design should be easier than for an RCT.

Retrospective Cohort Study

- Advantages of study design for producing a valid result: Significant risk of selection bias exists, and there is less ability to control for confounders than in a prospective cohort study because key variables may not be collected. However, this design could be sufficient for hypothesis generation that could then be used to design a more focused RCT.
- Resource use, size, and duration: A retrospective cohort study design has the potential to be considerably faster and less expensive than either an RCT or a prospective cohort study.
- Ethical, legal, and social issues: Confidentiality and Health Insurance Portability and Account (HIPAA)[†] issues may arise when diverse databases are linked without specific patient consent.
- Availability of data or ability to recruit: Recruiting is very feasible; the main concern is selection bias, depending on the source of the secondary data, and missing variables. Negotiations with the holders of the secondary data may take significant time.

Example 2: Table

Content area: Elective Cesarean section compared with planned vaginal delivery in healthy women. "What is the comparative effectiveness of planned Cesarean delivery versus planned vaginal delivery on maternal and neonatal outcomes?"[9]

[†] The Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy and Security Rules

Table 2. Comparison of study designs

Study Design Considerations	Randomized Trial	Nonrandomized Trial	Prospective Cohort
Description of design	Individual patients randomly assigned to planned vaginal or planned Cesarean at a predetermined period before delivery.	Individual practices assigned to protocols promoting a high use of planned Cesarean versus a low use of planned Cesarean delivery.	Individuals allowed to select either planned Cesarean or planned vaginal delivery; data collected on health status and delivery intent at a predetermined period before delivery, and continued data collection on changes in intent up to and through delivery.
Advantages of study design for producing a valid result	Although the design, if feasible, is likely to produce the most valid results, generalizability of a sample willing to be randomized on an issue imbued with personal preference and confounded by preexisting health factors is likely to be low.	The results of assigning practices will increase the generalizability of the results greatly but will sharply reduce the validity of the results. Another potential constraint to validity of results from nonrandom assignment of practices is that practices may cross over to greater or lesser promotion of planned Cesarean(s) over time.	Strong potential for confounding, but repeated measures of delivery intent may increase ability to control for confounding by indication. Good generalizability.
Resource use, size, and duration	Likely to require substantial resources to recruit sample large enough to evaluate rare neonatal outcomes.	Lower resource use than randomized studies but will require resources to ensure monitoring of fidelity to protocol. Recruitment of sufficient number of practices can be a constraint, so sample size considerations will continue to be an issue.	Low resource use other than data collection of intent and outcomes. Unlikely to have significant constraints on sample size.
Ethical, legal, and social issues	Significant ethical and legal issues with randomization.	Fewer ethical and legal issues in recruiting and assigning practices to a protocol than individuals.	No major ethical or legal issues.
Availability of data or ability to recruit	Poor.	Better than randomized trial, but challenges remain.	No major challenges to recruitment.

Example 3: Process or Methods Considerations

Content area: Treatment of prostate carcinoma. "Facilitate future research on potential biomarkers to identity patients whose disease is likely to be aggressive."[10]

Context: Although many efforts have been made to predict which patients with localized prostate cancer have aggressive disease, existing tools are inadequate to predict which patient to treat with any high degree of accuracy. With the emergence of biomarkers in other diseases, such as breast cancer, that have both prognostic and predictive power, the search continues to identify biomarkers that can predict which patients with prostate cancer face a poorer prognosis and may benefit to a greater degree from immediate treatment. Although a number of biomarkers have been explored to date with limited success, biomarkers continue to have a potentially important role.

Proposed research design: Establish biospecimen repositories with clinical data on diagnosis, treatment, and follow-up.

Study design considerations:

- Advantages of study design for producing a valid result*:* Biospecimen repositories create the resources needed to test the use of novel biomarkers in the future, while providing long-term data on outcomes that would take a long time to collect. Such repositories are being established for other studies, such as the ProtecT trial in the United Kingdom. In addition, given the differences in treatment regimens, populations, and possibly outcomes across studies, biospecimens from different trials might help address alternative hypotheses. The National Cancer Institute is establishing methods for each step of the process for creating and maintaining biospecimen repositories.
- Resource use, size, and duration: Although expensive to create and maintain, additional repositories will allow more biomarker testing, particularly because tissue specimens are finite. The administrative complexity of tracking specimens and their use is substantial, and ongoing infrastructural funding is essential.
- Ability to recruit: At the time of biopsy or surgery, patients could be consented for participation. Given that tissue is obtained as part of the procedure, this should be straightforward.
- Ethical, legal, and social issues: Biorepositories require extensive documentation of their policies regarding tracking and use of specimens. The proposed revisions to the Federal Department of Health and Human Services (HHS) Common Rule may partially address these issues.[‡] Significant planning will be needed.

Example 4: Sample Size Calculations To Inform Feasibility of Future RCTs

Content area: Coronary artery stenting compared with coronary artery bypass surgery.

Percutaneous coronary interventions (PCI) with or without stents and coronary artery bypass graft surgery (CABG) are clinically relevant treatment options for many patients with coronary

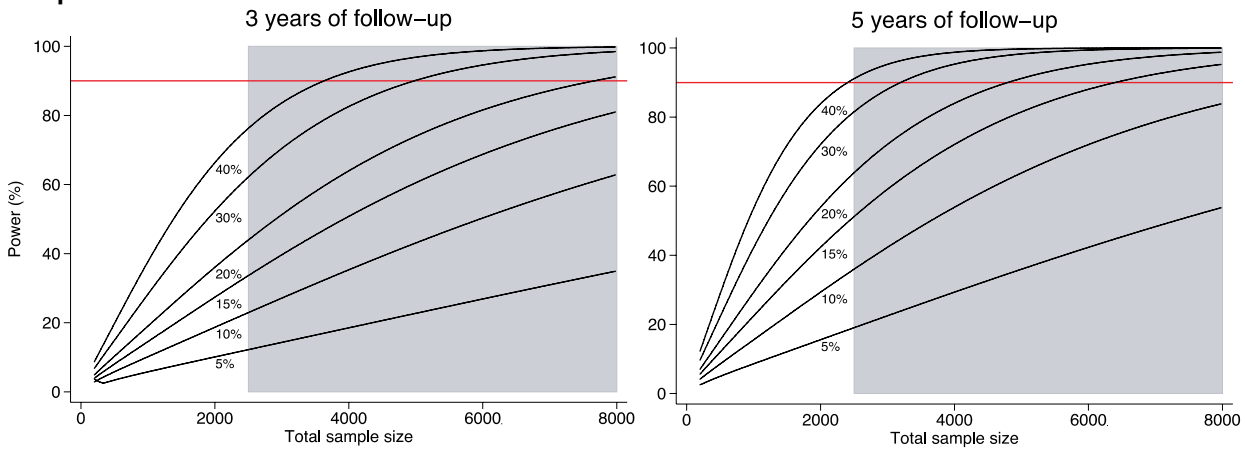
[‡] The Common Rule is located at 45 CFR part 46, subpart A, and described at: http://www.hhs.gov/ohrp/humansubjects/commonrule/index.html.

artery disease (CAD). In assessing this topic, it was deemed that an important gap pertained to the comparative effectiveness and safety of the interventions in the elderly (aged 75 or older).[11]

But exactly what merits further study? A focused value-of-information analysis helped clarify the group of parameters that would most inform the decision of choosing between PCI and CABG in the elderly. The analysis suggested that the *relative safety of the interventions* (i.e., relative effects on post-intervention complications) was more important than, for example, the absolute frequency of adverse events in the postintervention. As per descriptions above, a possible design to address relative effects of treatments is an RCT.

Is it realistic to consider a new RCT to compare PCI versus CABG? One can perform high-level sample size calculations. The biggest trials in the field enrolled approximately 2,500 patients, which serves as an indication of a large feasible RCT. Figure 2 shows power attained over a range of sample sizes for various control rate values over a mean followup of 3 or 5 years (see legend for details). Power increases with sample size, with control rate, and with length of followup. Over 5 years of followup, a study of approximately 2,500 patients would attain 80 to 90 percent power to find a relative effect of 0.80 only if it chooses an outcome that has at least a 30 percent control rate. This means a *composite outcome*. To get average followup duration of approximately 5 years, a trial would have to go on for 6 to 8 years at least (see legend).

Figure 2. Power calculations for superiority RCTs for various 5-year primary event rates in the comparator arm[11]



Plotted are power calculations for six different 5-year primary event rates in the comparator arm (5%, 10%, 15%, 20%, 30%, and 40%, as shown next to each line in each panel). The calculations are for a two-sided chi-squared test at the 0.05 level of significance, and assuming a constant annual event rate, a true relative effect of 0.80 favoring the intervention arm, an allocation ratio of 1:1, no loss to followup, no crossover between treatments, and no sequential monitoring. The gray area denotes sample sizes larger than the biggest existing RCTs on this question (>2,500 patients total). The red horizontal line stands for 90% power. Note that to get average followup duration of approximately 5 years, a trial would have to go on for 6 to 8 years. We calculate this assuming a minimum followup of 2.5 years, a patient recruitment period of 5 years, and a constant recruitment rate. In reality, the total sample size would have to be even larger than what is shown in the horizontal axis, as there will be loss to followup and there may also be adjustments for sequential monitoring.

Therefore, *de novo* RCTs are feasible but would likely require resources comparable to recent large RCTs. The above calculations are generic and thus apply to any subset of patients with heart disease. For example, in middle-aged patients with two vessel disease, one would have to define a composite outcome of death or myocardial infarction or other cardiac events to attain a high event control rate and, thus, high power to detect a significant difference. By contrast, in the subpopulation of elderly patients (e.g., older than 75 years), where mortality rates

can be high enough, one may be able to attain high statistical power for the outcome of death alone.

Conclusions

The incorporation of study design considerations into FRN documents provides an important addition to the priority-setting function of the reports. The summary guidance from the workgroup follows:

- Use of common terminology for study designs can aid in communication regarding FRNs. Such terminology can, of course, be modified when appropriate.
- Each high-priority research need may be addressed with one or more than one potential study design.
- EPC teams should incorporate brief discussions of study design considerations for the highest priority research needs into all FRN reports.
- Considerations for recommendation of a study design include ability of the study design to produce a valid result, as well as feasibility considerations such as resources needed; availability of data or ability to recruit and retain subjects; and ethical, legal, and social considerations.
- The PICOTS format can assist in making aspects of the FRN and the study design explicit but may not be appropriate for some types of study designs, such as methods enhancements.
- Some types of study designs that may be proposed, such as methods development, may not have been systematically reviewed in the original work, and prior to initiating the research project, investigators should determine whether the new work is duplicative.
- Sample size or power calculations may be helpful in considering the feasibility of future studies.

References

1. Institute of Medicine. Initial national priorities for comparative effectiveness research. Washington, DC: The National Academies Press; 2009.
2. Whitlock EP, Lopez SA, Chang S, et al. 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. PMID: 19540721.
3. Eden J, Levit L, Berg A, et al., eds. Finding what works in health care: standards for systematic reviews. Washington, DC: The National Academies Press; 2011.
4. Trikalinos T, Dahabreh I, Lee J, et al. Methods research on future research needs: defining an optimal format for presenting research needs. Methods Future Research Needs Report No. 3. Prepared by Tufts Evidence-based Practice Center under Contract No. 290-2007-10057-I. AHRQ Publication No. 11-EHC027-EF. Rockville, MD: Agency for Healthcare Research and Quality: 2011. www.effectivehealthcare.ahrq.gov/reports/final.cfm
5. Robinson KA, Saldanha IJ, Mckoy NA. Frameworks for determining research gaps during systematic reviews. Methods Future Research Needs Report No. 2. Prepared by Johns Hopkins University Evidence-based Practice Center under Contract No. 290-2007-10061-I. AHRQ Publication No. 11-EHC043-EF. Rockville, MD: Agency for Healthcare Research and Quality; June 2011. www.effectivehealthcare.ahrq.gov/reports/final.cfm
6. Owens DK, Lohr KN, Atkins D, et al. AHRQ Series Paper 5: Grading the strength of a body of evidence when comparing medical interventions: AHRQ and the Effective Health Care Program. J Clin Epidemiol. 2010;63:513-23. PMID: 21433399.
7. Hartling L, Bond K, Harvey K, et al. Developing and Testing a Tool for the Classification of Study Designs in Systematic Reviews of Interventions and Exposures. Prepared by University of Alberta Evidence-based Practice Center under Contract No. 290-02-0023. AHRQ Publication No. 11-EHC-007. Rockville, MD: Agency for Healthcare Research and Quality; December 2010. http://effectivehealthcare.ahrq.gov/
8. Butler M, Forte ML, Kane RL, et al. Future research needs for the treatment of common hip fractures. Future Research Needs Paper No. 2. Prepared by Minnesota Evidence-based Practice Center under Contract No. 290-2007-10064-I. AHRQ Publication No. 10-EHC071-EF. Rockville, MD: Agency for Healthcare Research and Quality; 2010. www.effectivehealthcare.ahrq.gov/reports/final.cfm
9. Viswanathan M, Visco AG, Hartmann K, et al. Cesarean Delivery on Maternal Request. Evidence Report/Technology Assessment No. 133. Prepared by RTI International-University of North Carolina Evidence-Based Practice Center under Contract No. 290-02-0016. AHRQ Publication No. 06-E009. Rockville, MD: Agency for Healthcare Research and Quality; 2006.
10. Mauger Rothenberg B, Marbella A, Belinson SE, et al. Future Research Needs for Comparative Effectiveness of Treatments for Localized Prostate Cancer. Future Research Needs Paper No. 4. Prepared by Blue Cross and Blue Shield Association, Technology Evaluation Center Evidence-based Practice Center under Contract No. 290-2007-10058-I. Rockville, MD: Agency for Healthcare Research and Quality; 2010. www.effectivehealthcare.ahrq.gov/reports/final.cfm

11. Trikalinos TA, Dahabreh IJ, Wong J, et al. Future Research Needs for the Comparison of Percutaneous Coronary Interventions with Bypass Graft Surgery in Nonacute Coronary Artery Disease: Identification of Future Research Needs from Comparative Effectiveness Review No. 9. Prepared by the Tufts Evidence-based Practice Center under Contract No. 290-2007-10055-I. AHRQ Publication No. 10-EHC068-EF. Rockville, MD: Agency for Healthcare Research and Quality; December 2010. www.effectivehealthcare.ahrq.gov/reports/final.cfm

12. Spaulding C, Daemen J, Boersma E, et al. A pooled analysis of data comparing sirolimus-eluting stents with bare-metal stents. N Engl J Med. 2007 Mar 8;356(10):989-97. PMID: 17296825.

13. Berger JS, Elliott L, Gallup D, et al. Sex differences in mortality following acute coronary syndromes. JAMA. 2009 Aug 26;302(8):874-82. PMID: 19706861.

Appendix A. Taxonomy for Study Designs[4]

The initial descriptive text following each design name comes from the Alberta report,[7] the subsequent text regarding design validity, resource use, recruitment and available data, and ethical, legal, and social issues is from the authors of the current report. The points raised for each dimension are intended as a starting point to inspire discussion, rather than a set recipe for evaluation.

A. Randomized Controlled Trial (RCT)

An RCT is a study designed to test the efficacy of an intervention on an individual, a group of individuals, or clusters (e.g., classrooms, communities). Individuals or clusters are randomly allocated to receive an intervention or control/comparison (e.g., placebo or another intervention) and are followed prospectively to assess differences in outcomes. The unit of analysis is the individual, group of individuals, or the cluster, as appropriate. Many CER interventions addressed by EPC's involve interventions at the level of the provider or the practice through which patients are clustered within another unit. Randomization among a relatively small number of clusters requires planning so as to address the need for balance in baseline characteristics and address sample size considerations in the clustered design. Variations in treatment assignment and measurement produce different types of studies including factorial, crossover, parallel, stepped wedge and Solomon four-group, and adaptive designs.[7]

Advantages of study design for producing a valid result: An RCT is the best method to control for selection bias and both measured and unmeasured differences between groups at baseline, but these benefits may come at the cost of generalizability. Pragmatic trial strategies may be appropriate to address this challenge in some cases. At the same time, the RCTs of some questions (particularly long-term outcomes) may be vulnerable to crossover, leading to problems in interpretation of "intention to treat" analyses.

Resource use, size, and duration: Generally, RCTs are high cost in dollars and investigator time, although some types of trial designs, which involve less data collection from subjects than other types of designs, may reduce the resources needed. Study duration will depend on the underlying condition assessed. For example, duration of acute infectious diseases may require only a brief followup. As these are prospective studies, duration is longer than retrospective studies. Study size will depend on the effect size being sought between intervention and comparator groups.

Availability of data or ability to recruit: This factor varies by topic. Even if the research community believes that equipoise exists, patients may be reluctant to accept randomization for some interventions or medical conditions.

Ethical, legal, and social issues: Concerns may occur when treatment is assigned though random allocation. Patients and treating clinicians must perceive equipoise across tests or treatments when invited to participate in a research study. Careful stopping and reporting rules will be important if evidence of significant benefit or harm is found. RCTs are typically performed when there is equipoise on the optimal treatment. Equipoise is topic-specific: most often, it refers to expected patient health-relevant outcomes (clinical equipoise), but it may also refer to optimal resource allocation and other considerations.

[4] Citations in this appendix are to the report's main reference list.

B. Nonrandomized Comparative Trial

A nonrandomized comparative trial is a study in which individuals or groups of individuals (e.g., community, classroom) are assigned to the intervention or control by a method that is not random (e.g., date of birth, date of admission, judgment of the investigator). Individuals or groups are followed prospectively to assess differences in the outcome(s) of interest. The unit of analysis is the individual or the group, as appropriate.[7]

Advantages of study design for producing a valid result: Selection bias by subjects will be partially controlled for since the researchers will assign treatment, although the groups may not be balanced at baseline. Prospective design allows better assessment of baseline status. This design might be appropriate when allocation is 'clustered' at the unit of intervention, such as practices, health care facilities or geographic areas. In such situations, purposive sampling may be the best way to assure balancing of covariates between intervention and comparator groups. Some rules for assignment, such as birth date or date of clinical visit, may be susceptible to 'gaming.' When study duration is long, crossover may occur, leading to difficulty in interpretation according to 'intention to treat' analyses.

Resource use, size, and duration: Resources needed are generally high, but can be less expensive than randomized trials. Study duration will depend on the underlying condition being assessed. For example, duration of acute infectious diseases may require only a brief follow-up. Because these are prospective studies, duration is longer than retrospective studies. Size will depend on the effect size being sought between intervention and comparator groups.

Availability of data or ability to recruit: Acceptability to potential subjects is better than RCTs in which a group is assigned at random. One disadvantage is that subjects may be less willing to accept an assignment as opposed to maintaining choice.

Ethical, legal, and social issues: Concerns may occur when treatment is assigned by investigators. Patients and treating clinicians must perceive equipoise across tests or treatments when invited to participate in a research study. Careful stopping and reporting rules will be important if evidence of significant benefit or harm is found.

C. Prospective Cohort Study

In a prospective cohort study, individuals in the group without the outcome(s) of interest (e.g., disease) are classified according to exposure status at baseline (exposed or unexposed) and then are followed over time to determine if the development of the outcome of interest is different in the exposed and unexposed groups. [7]

Advantages of study design for producing a valid result: This type of study may be the optimal design when treatment or exposure cannot ethically or practically be assigned. Baseline characteristics can be measured, but may not be balanced between the two groups. If sample size is large, this design may be the best method to assess subgroup effect or incidence of harms. Statistical techniques to adjust for baseline differences may not completely control for potential bias. Crossover may occur, especially in longer duration observation.

*Resource use, size, and duration***:** Resources needed are moderate to high; similar to a prospective trial, although less than a study design in which treatment is assigned. Size and duration will depend on the natural history of the condition under study, and the effect size or incidence of harm thought to be clinically important.

Availability of data or ability to recruit: Acceptability to potential subjects is better than designs in which treatment is assigned. The ability to recruit may depend on the respondent burden.

Ethical, legal, and social issues: Since patients and/or providers select the treatment, few ethical issues are likely to occur, although careful stopping and reporting rules will be important if evidence of significant benefit or harm is found.

D. Retrospective Cohort Study

In a retrospective cohort study, a group of individuals is identified by common features that were determined in the past. The group is usually assembled using available data sources (e.g., administrative data). Individuals are classified according to exposure status (exposed or unexposed) at the time the group existed and are followed up to a prespecified endpoint to determine if the development of the outcome of interest is different in the exposed or unexposed groups.[7]

Advantages of study design for producing a valid result: Subjects are followed over time. The duration of the study is short since it is retrospective. Baseline measures are not prospectively assessed; the researchers must accept the data already collected, so key variables may not be added. Significant risk of selection bias exists, since subjects either choose their treatment or it is assigned by a health care provider. The generalizability of the study result will depend on the population sampled.

Resource use, size, and duration: Since these studies often use data already collected, they are much less resource intensive than studies employing prospective data collection. The duration of the study is short relative to studies involving primary data collection. However, such studies often take somewhat longer than envisioned. Sample size may be very large given the increasing availability of large administrative claims and electronic medical record (EMR) databases.

Availability of data or ability to recruit: Recruiting is very feasible; the main concern is selection bias, depending on the source of the secondary data, and missing variables. Negotiations with the holders of the secondary data may take significant time.

Ethical, legal, and social issues: Confidentiality and HIPAA[5] issues may arise when diverse databases are linked without specific patient consent.

E. Interrupted Time Series With Comparison Group

In this type of study, multiple observations over time are "interrupted" by an intervention or exposure, and two series are examined (one is a comparison group). At least three observations before and at least three observations after the intervention or exposure must take place for each group. The investigator(s) does not assign or have control over the intervention/exposure, which may be an environmental variable (e.g., airborne toxin) or administrative assignment (e.g., seatbelt legislation, educational program, service delivery model) but does control the timing of the measurement and the variables being measured.[7]

Advantages of study design for producing a valid result: This type of study may be the optimal method for evaluating some policy-level interventions (such as legislation or an insurance policy change) for which assignment cannot practically be performed. There may be differences in both measured and unmeasured characteristics at baseline. Other co-interventions may occur in the health care environment, leading to difficulties in attribution of the observed change in outcome to the intervention of interest. However, the investigator may not have any control over the number or timing of observation points.

[5]The Health Insurance Portability and Accountability Act of 1996 (HIPAA) Privacy and Security Rules

Resource use, size, and duration: Resources needed are generally modest, depending on the amount of new data that needs to be collected. The duration of the study is short relative to studies involving primary data collection. However, such studies may often take longer than planned. Sample size can be very large given the increasing availability of relatively inexpensive large administrative claims and EMR databases.

Availability of data or ability to recruit: Recruiting is very feasible, since the study is conducted as a naturally occurring experiment. Negotiations with the holders of the secondary data may take significant time.

Ethical, legal, and social issues: Such issues should be minimal, although some issues regarding HIPAA and data use agreements may be relevant.

F. Controlled Before–After Study

In a controlled before-after study, the outcome(s) of interest is measured both before and after the intervention or exposure in two or more groups of individuals. In this study design the study group receives the intervention or exposure and the comparison group(s) does not. This type of study includes interventions that may be in the control of the investigator (e.g., a surgical procedure) as well as interventions that may be an environmental variable (e.g., airborne toxin) or administrative assignment (e.g., seatbelt legislation). In all cases the investigator(s) controls the timing of the measurement and the variables being measured.[7]

Advantages of study design for producing a valid result: This type of study may provide greater validity than an interrupted time series study, which is purely observational. This study design may provide better assessment of baseline variables for assessment of confounding. The study groups are likely to be unbalanced at baseline, and statistical measures to control for differences in group characteristics are likely to be only partially successful. Residual confounding will likely be present.

Resource use, size, and duration: Resources needed are somewhat greater than for interrupted time series design, since consent will be required for many interventions. Size of the study sample may range from small to very large. Small studies may involve one or a few providers or practices and may approximate case studies. Duration will depend on the outcomes selected and the anticipated lag between the intervention and the expected clinical outcome.

Availability of data or ability to recruit: Recruiting is fairly feasible, but some subjects may refuse to participate. This design may be the best way to study some interventions where randomization may not be acceptable to providers or patients.

Ethical, legal, and social issues: If the providers or patients are not at equipoise, ethical issues may be pertinent if individuals are perceived as not receiving efficacious care.

G. Nonconcurrent Cohort Study

In a nonconcurrent cohort study, two or more groups of individuals are identified on the basis of common features at different time points. Individuals in each group are classified according to exposure status (exposed or unexposed) at the time the groups existed or were created. They are followed to determine if the development of the outcome of interest is different in the exposed or unexposed groups.[7]

Advantages of study design for producing a valid result: Selection bias and group differences in patient characteristics may lead to residual confounding. The 'historical controls' may have differences in either baseline characteristics, or, commonly, differences in co-interventions which may be difficult to adjust for statistically.

Resource use, size, and duration: The resources needed are relatively low compared with randomized or controlled trial designs, but greater than entirely retrospective studies. The study size may range from small to very large, depending on the research question and the availability of data. Duration will depend on the outcomes selected and the anticipated lag between the intervention and the expected clinical outcome.

Availability of data or ability to recruit: Good. Patients and providers choose the treatments so acceptance will be high. New data collection may or may not be needed. This study design generally uses existing data; negotiation with the owner of the data regarding access may take time.

Ethical, legal, and social issues: Such issues should be minimal, since the exposure is not being assigned by the investigator. The usual ethical issues of informing subjects of potential benefits and harms will apply.

H. Nested Case Control Study

In a nested case control study, exposed and control subjects are drawn from the population of a prospective cohort study. Baseline data are obtained at the time the population is identified; the population is then followed over a period of time. The study is then carried out using persons in whom the disease or outcome has developed and a sample of those who have not developed the outcome of interest (controls).[7]

Advantages of study design for producing a valid result: This study design is most appropriate for etiologic studies. It uses existing data, and is relatively quick to conduct. It has some advantages over the usual case-control study, since the assessment of subject characteristics is generally reliable within the original cohort study, which may use primary data collection. The case control study design does, however, have multiple threats to causal inference. The original cohort should be assessed regarding its representativeness for the population of interest.

Resource use, size, and duration: Resources needed are low. The size of the study is generally modest, since it is nested within an existing cohort or other prospective study. The duration of the study is relatively brief since the investigators are generally working from previously collected data.

Availability of data or ability to recruit: High, since researchers are often sampling using already-collected data.

Ethical, legal, and social issues: Such issues should be minimal, since the study is observational.

I. Case Control Study

In a case control study, participants are selected based on the known outcome(s) of interest (e.g., disease, injury). Exposure status is then collected based on the participants' past experiences. Exposure status is compared between the two (or more) groups: those who have the outcome of interest and those who do not have the outcome of interest (controls). This is a retrospective study that collects data on events that have already occurred.[7]

Advantages of study design for producing a valid result: This study design is most appropriate for etiologic studies, and is relatively quick to conduct. The case control study does have multiple threats to causal inference inherent to the study design. Choice of controls as representing those at risk for the outcome of interest is key to validity.

Resource use, size, and duration: Few resources are needed. The size of case control studies is variable, and in part depends on the number of controls selected for each case. Sample sizes are generally modest. The duration is relatively brief since the study design is retrospective.

Availability of data or ability to recruit: High, since researchers are often sampling from already collected data.

Ethical, legal, and social issues: Such issues should be minimal, since the study is observational.

J. Interrupted Time Series (Without a Comparison Group)

In this type of study, multiple observations over time are "interrupted" by an intervention or exposure. There must be at least three observations before the intervention and at least three observations after the intervention; otherwise, the study is considered a before-after study. The investigator(s) does not assign or have control over the intervention/exposure, which may be an environmental variable (e.g., airborne toxin) or administrative assignment (e.g., seatbelt legislation, educational program, service delivery model) but does have control over the timing of the measurement and the variables being measured.[7]

Advantages of study design for producing a valid result: This study design involves observation of a natural experiment. The changes observed may be difficult to differentiate from secular trend in occurrence of the outcome due to factors other than the variable of interest. No comparison group is present, limiting its utility in CER.

Resource use, size, and duration: Resources needed are relatively low. Study size can be small to large, depending on the intervention and the availability of data. Duration of the study will depend on the expected accrual of outcome events and the lag between the intervention and outcomes.

Availability of data or ability to recruit: High; generally uses already collected administrative or, in the future, EMR data.

Ethical, legal, and social issues: Such issues should be minimal, since the study is observational. There may be issues regarding data use agreements or HIPAA issues, but data can generally be appropriately de-identified to minimize risk of inappropriate disclosure.

K. Before–After Study

This is a study of an intervention or exposure in which the investigator(s) compares the outcome(s) of interest both before and after the intervention in the same group of individuals. This includes interventions that may be in the control of the investigator (e.g., a surgical procedure) as well as interventions that may be an environmental variable (e.g., airborne toxin) or administrative assignment (e.g., seatbelt legislation). In all cases the investigator(s) controls the timing of the measurement and the variables being measured.[7]

Advantages of study design for producing a valid result: This study has a simple design, similar to a case series study of an intervention. It occurs in a naturalistic setting and may have high generalizability. It may be difficult or impossible to attribute causal inference to the intervention since there is no comparison group. Natural history of the condition may be unknown.

Resource use, size, and duration: Resources needed are generally low, but depend on the cost of measurement. Study size may be small or large, depending on the intervention or the availability of data. Duration of the study will depend on the expected accrual of outcome events and the lag between the intervention and outcomes.

Availability of data or ability to recruit: High.

Ethical, legal, and social issues: Such issues should be minimal, since patient and provider select the treatment.

L. Cross-Sectional Study

In a cross-sectional study, both the exposure and the outcome status in a target population are assessed concurrently, that is, at the same point in time or during a brief period of time. The temporal sequence of cause and effect cannot necessarily be determined. They are most commonly used to assess prevalence. Data collection is commonly done by using a survey.[7]

Advantages of study design for producing a valid result: This design may be highly generalizable to the population of interest, but one cannot infer causation to an intervention.

Resource use, size, and duration: Moderate resources are needed, but significant effort may be required to generate generalizable samples of the population. The sample size may range from small to large, depending on the study question and the expected incidence of the condition or care utilization being studied.

Availability of data or ability to recruit: High, although response rates on surveys have been lower recently due to use of cell phones and caller identification.

Ethical, legal, and social issues: Such issues should be minimal to none; no intervention is involved. Confidentiality may be an issue for sensitive issues (HIV, mental health, etc.).

M. Noncomparative Study

Examples of this design include:

- A study that presents a description of a single patient or participant. Studies are usually retrospective and typically describe the manifestations, clinical course, and prognosis of the individual.
- A study that describes the experience of a group of patients with a similar diagnosis and/or treatment. Studies are usually retrospective and typically describe the manifestations, clinical course, and prognosis of a condition.
- A study in which data are collected at a series of points in time on the same population to observe trends in the outcome(s) of interest.[7]

Advantages of study design for producing a valid result: This type of design may describe a new or unique condition or intervention. The study is hypothesis generating rather than testing. Generally, one is unable to infer cause. Generalizability may be poor if the patients or setting are different from those in which patients with the condition of interest are usually seen. No comparison group is involved, making the utility of the study design in CER difficult.

Resource use, size, and duration: Resources needed are low. The sample size is generally small, since the study involves description of single or several subjects. Study duration is short and the data collection is often retrospective.

Availability of data or ability to recruit: Good, generally derived from existing medical record or administrative data.

Ethical, legal, and social issues: Such issues should be minimal; treatment is selected by patients and providers.

N. Additional Considerations

Meta-analysis of individual participant data, modeling, and additional systematic reviews of group data are additional analyses that do not necessitate collection of additional primary data.

They can be reasonable choices for addressing FRNs, for informing choices between possible future studies, or for informing the planning of a definitive future study.

Meta-analysis of individual participant data (MIPD). In a meta-analysis of individual participant data (MIPD), data from existing studies are brought together using harmonized definitions, and reanalyzed according to a prespecified protocol. When individual participant data are available, outcome and exposure definitions can be standardized and much more powerful analyses can be used. A major advantage over a meta-analysis of group data is the ability to examine patient-level modifiers of the treatment effect, i.e., patient-level factors such as age, sex, and disease severity indices.

An MIPD can take longer to complete than an analysis of a readily available dataset from a single study. Logistical complications include, but are not limited to, identification of data sources, convincing investigators to participate, standardizing definitions of interventions and outcomes, complying with HIPAA, and harmonizing definitions of exposures and outcomes across datasets.

We maintain that if an MIPD is feasible, it should be considered as a possible method for studying questions on clinical heterogeneity in a timely manner. There are several examples of MIPDs that have generated answers to important questions.[12, 13]

Modeling. Modeling includes quantitative decision and economic (cost-effectiveness, cost-utility, and value of information) analysis. Many evidence gaps can be reasonably explored with modeling approaches. An example is the assessment of screening programs with respect to clinical outcomes. Even if one considers a single test, e.g., screening colonoscopy for preventing colorectal cancer, there are many choices regarding the age to start screening, how often to screen, and the age to stop screening. Studying a large number of screening schedules is impossible in an RCT—it may be difficult to study even two. Modeling can help explore which schedules are completely suboptimal and do not merit further study. It can also assist in determining which screening schedules are expected to be better, given the currently available information. Indeed, the United States Preventive Services Task Force took into account insights gained from modeling in formulating their screening recommendations for colorectal and breast cancer. Scrutiny at the level of comparisons among many screening schedules and with no screening is impossible without simulation modeling.

Additional Systematic Reviews. Evidence gaps and FRNs are derived from a systematic review, of course, and it may seem redundant that a systematic review would be recommended as a research need derived from a systematic review. However, a subanalysis that may have been out of scope of the original review, or derived from the review key questions but not assessed in the parent systematic review, could be considered an evidence gap. However, if the proposed evidence gap is conceptually quite far from the original key questions covered in the parent systematic review, then the scope of the FRN gaps identification would expand dramatically and potentially lose focus.

O. Additional Concepts

The terms in Table A-1 are derived from the Alberta report, and are commonly used in association with the study designs discussed above.[7]

Table A-1. Terms associated with study designs

Term	Description
Cluster	The term cluster refers to a unit of allocation or analysis in a clinical trial. Examples of clusters include hospitals, schools, neighborhoods, or entire communities.
Cluster randomized controlled trials	Synonym: *community trial; group randomized trial.* A randomized controlled trial in which the units of randomization and analysis are groups of people or communities (e.g., classroom, hospital, town). Typically, several communities receive the intervention and several different communities serve as controls.
Cohort	The term cohort refers to a group of individuals (or other organizational units) who have a common feature when they are assembled (e.g., birth year, place of employment, medical condition, place or time period of medical treatment) and are followed over time. They can be followed prospectively or examined retrospectively.
Experimental study	A type of study in which investigators directly control the timing, course, and assignment of the intervention. Experimental studies investigate an intervention to determine its effect on the outcome(s) of interest. In an experimental study a population is selected to receive a specific intervention, the effects of which are measured by comparing the outcomes in the experimental group with the outcomes of a control group that has received another intervention or placebo. Examples include randomized controlled trial, cluster randomized controlled trial, nonrandomized trial, n-of-one trial.
Observational study	A study in which the investigator(s) does not control the exposure/ intervention status of study participants (i.e., the assignment of the intervention or exposure of interest is not under the control of the investigator(s)). The simplest form of observational study is the case report or case series, which describes the clinical course of individuals with a particular condition or diagnosis. Observational studies include descriptive and analytic studies.
Quasi-experimental study	A type of study in which the investigator(s) evaluates the effect of an intervention but does not fully control the timing, course, or allocation of the intervention. They are often used when it is not possible to conduct a true experimental study.