Abstract

Objective: This article describes some of the fundamental principles that have been developed to guide the work of producing comparative effectiveness reviews (CERs).

Study Design and Setting: We briefly describe the role stakeholders play in providing important insights that inform the evidence-gathering process, and discuss the critical role of analytic frameworks in illuminating the relationship between surrogate measures and health outcomes, providing an understanding of the context in which clinical decisions are made and the uncertainties that underlie clinical controversies.

Results: We describe the Effective Health Care program conceptual model for considering different types of evidence that emphasizes minimizing the risk of bias, but places high-quality, highly applicable evidence about effectiveness at the top of the hierarchy. Finally, we briefly describe areas of future methodological research.

Conclusion: CERs have become a foundation for decision-making in clinical practice and health policy. To be useful, CERs must approach the evidence from a patient-centered perspective; explore the clinical logic underlying the rationale for a service; cast a broad net with respect to types of evidence, placing a high value on effectiveness and applicability, in addition to internal validity; and, present benefits and harms for treatments and tests in a consistent way. © 2010 Elsevier Inc. All rights reserved.

Keywords: Comparative effectiveness reviews; Stakeholders; Hierarchy of evidence; Patient-centered care; Avoiding bias; Evidence-based medicine

1. Introduction

Comparative effectiveness reviews (CERs) are summaries of available scientific evidence in which investigators collect, evaluate, and synthesize studies in accordance with an organized, structured, explicit, and transparent methodology. They seek to provide decision-makers with accurate, independent, scientifically rigorous information for comparing the effectiveness and safety of alternative clinical options, and have become a foundation for decision-making in clinical practice and health policy. To play this important role in decision-making, CERs must address significant questions that are relevant to patients and clinicians, and use valid, objective, and scientifically rigorous methods to identify and synthesize evidence and apply them consistently and in an unbiased and transparent manner.

In this article, we describe the preliminary work and key principles that underlie the development of the

Methods Guide for Comparative Effectiveness Reviews (http://effectivehealthcare.ahrq.gov/healthInfo.cfm?infotype=rr&ProcessID=60). The articles in this series describe recommended approaches for addressing difficult, frequently encountered methodological issues. Although the science of systematic reviews is evolving and dynamic, excessive variation in methods among systematic reviews gives the appearance of arbitrariness and idiosyncrasy that undercuts the goals of transparency and scientific impartiality.

2. Background and history

In 1997, the Agency for Healthcare Research and Quality (AHRQ) began its Evidence-based Practice Center (EPC) program. EPCs were established and staffed with personnel who had training and expertise in the conduct of systematic evidence reviews. From its inception, the EPCs have been committed to developing methods for identifying and synthesizing evidence that minimize bias, and adopted some precautions against bias in conducting evidence reviews that were extraordinary for their time.
In 1996, for example, the procedures used by EPCs, documented in the EPC Procedure Manual, included a requirement for the involvement of a technical expert panel to work with EPC scientists to develop the questions to be answered in the review as a way to protect against bias in framing or selecting questions [1]. This approach helps ensure that a review will address important questions that decision-makers need answered, and also protects against bias in framing or selecting questions. Another protection against reviewer bias—using independent researchers, without conflicts of interest, to assess studies for eligibility—has also been used since the inception of the EPC program.

The Guide is part of a broader system of safeguards to ensure that reviews produced by the EPCs are of high-quality, consistent, and fair [2]. Safeguards are needed because, as in any type of clinical research, the habits or views of investigators and funders can introduce bias, variation, or gaps in quality [3–5]. The framework for conducting systematic reviews includes strategies to reduce the possibility of bias at every step [6,7].

The Guide is a collaborative product of the 14 EPCs with oversight from the Scientific Resource Center (SRC) for the Effective Health Care Program and scientists employed by AHRQ. To prioritize topics for the Guide, we

- Identified challenges in the production of AHRQ evidence reports and variation among EPCs; and
- Examined public and peer-reviewed commentary on CERs.

In 2004 and 2005, each EPC analyzed published evidence reports and produced a series of articles identifying methodological challenges and areas of high practice variation among the EPCs. Topics included assessing beneficial [8] or harmful effects of interventions [9]; using observational studies [10]; assessing diagnostic tests [11] or therapeutic devices [12]; and others. When possible, the articles also suggested best practices [13].

Through these approaches, we have identified concerns about inconsistent or poorly developed methods that are common across reports such as:

- Inconsistency in approaches to quantitative synthesis such as the choice of fixed or random effects model;
- Inconsistency in the selection of data sources and evaluation of their quality for assessment of harms;
- A weakly developed approach to assessing the strength of evidence and a desire to begin to reconcile the EPC and Grading of Recommendations Assessment, Development and Evaluation (GRADE) approaches; and
- A need to develop a consistent and structured approach to the assessment of applicability.

We used this preliminary work to select the key issues for the first version of the Guide. To address these issues, AHRQ established five workgroups made up of EPC investigators, AHRQ staff, and SRC staff. The five workgroups developed guidance on observational studies, applicability, harms and adverse effects, quantitative synthesis, and methods for rating a body of evidence. The workgroups identified relevant methods articles and reviewed the published guidance from major bodies producing systematic reviews—most importantly the Cochrane Collaboration Handbook [14] and the Center for Reviews and Dissemination manual on conducting systematic reviews [15,16].

2.1. Principles for developing guidance

The fundamental principle used in the development of the Guide and subsequent guidance has been that workgroups should use empiric, methodological research when available. However, when empirical evidence is not available or is inadequate, workgroups are asked to develop a structural, best-practice approach based on the principle that the approach will eliminate or reduce variation in practice and provide a transparent and consistent methodological approach.

Searching databases of non-English language publications, unpublished articles, and information published only in abstract form is an example of evidence-based guidance based on empiric research. Many publications on these topics exist [17–19] and they form a cohesive and consistent body of evidence on which recommendations can be made.

On the other hand, centralization of activities at the SRC, such as searching clinical trial registries and the FDA Web site, where EPC proficiency and skills vary; adoption of strict policies regarding conflicts of interest; and the introduction of an editorial review process that provides for an independent judgment of the adequacy of an EPCs response to public and peer-review comments are examples of structural approaches designed to reduce variation in practice and assure consistency across EPCs. Some of the most important structural components of the Effective Health Care Program are intended to ensure that patients’ and clinicians’ perspectives are heard, by standardizing the governance of interactions with technical experts, stakeholders, and payers.

2.2. Principles for conducting CERs

In their charge, all workgroup participants were asked to make their guidance for conducting reviews consistent with the overarching principles of the Effective Health Care Program [20]. Principles for conducting reviews include:

1. Approaching the evidence from a clinical, patient-centered perspective;
2. Fully exploring the clinical logic underlying the rationale for a service;
3. Casting a broad net with respect to types of evidence, placing a high value on effectiveness and applicability, in addition to internal validity; and
4. Presenting benefits and harms for different treatments and tests in a consistent way so that decision-makers can fairly assess the important tradeoffs involved for different treatment or diagnostic strategies.

For example, to follow the principle of patient-centeredness, the program encourages EPCs to use absolute measures whenever possible to promote better communication with patients and others who will use the reports. Similarly, the program has been aggressive in involving stakeholders at every step of the process to ensure public participation and transparency [21].

The EPCs’ approach to evidence synthesis incorporates important insights from clinical epidemiology, health technology assessment, outcomes research, and the science of decision-making [22,23]. These principles for conducting reviews reflect the EPC program’s longstanding commitment to developing evidence reports that individuals and groups can use to make decisions and that are relevant and timely, objective and scientifically rigorous, and provide for public participation and transparency.

2.2.1. Clinical and patient-centered perspective

Whoever the intended users are, a CER should focus on patients’ concerns. As Black notes, “There is no inherent antithesis between patient-oriented medicine and evidence-based medicine; focus on what is perceived by the individual patient does not rule out a systematic search for evidence relevant to his treatment [24].” Patients’ preferences and patient-centered care are fundamental principles of evidence-based medicine [25]. These principles mean that, regardless of who nominates a topic and who might use CERs, the reviews should address the circumstances and outcomes that are important to patients and consumers. Studies that measure health outcomes (events or conditions that the patient can feel and report on, such as quality of life, functional status, or fractures) are emphasized over studies of intermediate outcomes (such as changes in blood pressure levels or bone density). They should also take into account that, for many outcomes and decisions, variation in patients’ values and preferences can and should influence decisions [26]. Interviews with patients and studies of patients’ preferences when they are available, are essential to identify pertinent clinical concerns that even expert health professionals may overlook [8]. AHRQ has developed explicit processes for topic selection and refinement and the development of key questions to ensure that CERs are both patient-centered and meet the needs of other stakeholders [21].

2.2.2. Clinical logic and analytic frameworks

An evidence model is a critical element for fully exploring the clinical logic underlying the rationale for a service [27]. In the EPC program, the most commonly used evidence model is the “analytic framework [28,29].” The analytic framework portrays relevant clinical concepts and the clinical logic underlying beliefs about the mechanism by which interventions may improve health outcomes [30]. In particular, the analytic framework illustrates and clarifies the relationship between surrogate or intermediate outcome measures (such as cholesterol levels) and health outcomes (such as myocardial infarctions or strokes) [31], and when properly constructed, can provide an understanding of the context in which clinical decisions are made and illuminate disagreements about the clinical logic that underlie clinical controversies.

An analytic framework can also help clarify implicit assumptions about benefits from healthcare interventions, including assumptions about long-term effects on quality of life, morbidity, and mortality. These assumptions often remain obscure without a framework that can lead technical experts and manufacturers of drugs and devices to make explicit the reasoning behind their clinical theories linking surrogate outcomes, pathophysiology, and other intermediate factors to outcomes of interest to patients, clinicians, and other health care decision-makers.

Figure 1 depicts an analytic framework for evaluating studies of a new enteral supplement to heal bedsores. Key questions are associated with the linkages (arrows) in the analytic frameworks. When available, evidence that directly links interventions to the most important health outcomes is more influential than evidence from other sources. In the figure, Arrow 1 corresponds to the question (Key question 1): Does enteral supplementation improve mortality and quality of life?

In the absence of evidence directly linking enteral supplementation with these outcomes, the case for using the nutritional supplement depends on a series of questions representing several bodies of evidence:

- Key question 2: Does enteral supplementation improve wound healing?
- Key question 3: How frequent and severe are side effects such as diarrhea?
- Key question 4: Is wound healing associated with improved survival and quality of life?

Note that in the absence of controlled studies demonstrating that using enteral supplements improves healing (link #2), EPCs may need to evaluate additional bodies of evidence. Specifically, this would include evidence linking enteral supplementation to improved nutritional status and other evidence linking nutritional status to wound healing. Studies that measure health outcomes directly are given more weight, but the analytic framework makes clear what surrogate outcomes may represent them, and what bodies of evidence link the surrogate outcomes to health outcomes.

2.2.3. Types of evidence

Historically, evidence-based medicine has been associated with a hierarchy of evidence that ranks randomized trials higher than other types of evidence in all possible situations [32,33]. In recent years, broader use of
systematic CERs has brought attention to the danger of overreliance on randomized clinical trials and suggestions for changing or expanding the hierarchy of evidence to take better account of evidence about adverse events and effectiveness in actual practice [34–36].

AHRQ’s EPC program from the outset has taken a broad view of eligible evidence [1,37]. AHRQ reviews published between 1997 and 2005 encompassed a wide variety of study designs, from randomized controlled trials (RCTs) to case reports. In contrast to Cochrane reviews, most of which exclude all types of evidence except for RCTs, inclusion of a wider variety of study designs has been the norm rather than the exception in the EPC program [9–11,27,38,39].

In the Effective Health Care Program, the conceptual model for considering different types of evidence still emphasizes minimizing the risk of bias, but it places high quality, highly applicable evidence about effectiveness at the top of the hierarchy. The model also emphasizes that simply distinguishing RCTs from observational studies is insufficient because different types of evidence except for RCTs vary in their usefulness in CERs.

Discussions about the role of nonrandomized studies often focus on the limitations of RCTs and invoke the distinction between effectiveness and efficacy. Efficacy trials (explanatory trials) determine whether an intervention produces the expected result under ideal circumstances. Efficacy studies use less stringent eligibility criteria, assess health outcomes, and have longer follow-up periods than most efficacy trials. Roughly speaking, effectiveness trials measure the degree of beneficial effect under “real world” clinical settings [40]. The results of effectiveness studies are more applicable to the spectrum of patients that will use a drug, have a test, or undergo a procedure than results from highly selected populations in efficacy studies. Characteristics of efficacy trials that limit the applicability of their results include:

- Homogeneous populations. Trials may exclude patients from important subpopulations or those with relevant comorbidities;
- Small sample size;
- Limited duration;
- Focus on intermediate or surrogate outcomes; and
- Selective focus on a limited number of intended or unintended effects.

In contrast, effectiveness studies aim to study patients who are likely to be offered the intervention in everyday practice. They also examine clinical strategies that are more representative of or likely to be replicated in practice. They may measure a broader set of benefits and harms (whether anticipated or unanticipated) including self-reported measures of quality of life or function [41] and long-term outcomes that require longitudinal data collection to measure.

When they are available, head-to-head effectiveness trials—randomized trials that meet the criteria for effectiveness studies—are the best evidence to assess comparative effectiveness. Effectiveness trials enable the investigator to obtain evidence about effectiveness while minimizing the risk of bias from confounding by indication and other threats to internal validity [40,42–47]. The ideal trial

- Has good applicability to the patients, comparisons, setting, and outcomes important to patients and clinicians;
- Has a low risk of bias;
- Directly compares interventions;
- Reflects the complexity of interventions in practice; and

![Fig. 1. Analytic framework for a new enteral supplement to heal bedsores.](image-url)
Includes all important intended and unintended effects, taking adherence and tolerability into account.

Often, RCTs are deficient in one or more of these respects. The decision to use other kinds of evidence—experimental or observational—should follow a critique of the applicability, risk of bias, directness, and completeness of the RCT evidence [10]. In addition to head-to-head effectiveness trials, types of evidence used in CERs include:

- Long-term head-to-head controlled trials focusing on a subset of relevant benefits or risks;
- Cohort, case-control, or before/after studies with broad applicability and comprehensive measurement of benefits and risks;
- Short-term head-to-head trials that use surrogate (efficacy) measures;
- Short-term head-to-head trials focusing on tolerability and side effects;
- Placebo-controlled trials demonstrating an important or unique benefit or harm of a particular drug;
- Before/after or time-series studies demonstrating an important or unique benefit or harm of a particular drug;
- Natural history (or conventionally treated history) studies that observe the outcomes of a cohort but do not compare the outcomes among different treatments; and
- Case series and case reports.

In any particular review, any or all of these types of studies might be included or rendered irrelevant by stronger study types. Usually the reasons to include them overlap—RCTs may have poor applicability because of patient selection or inappropriate comparator or dosing of comparator; may not address all relevant intended effects; may not address all relevant unintended effects; or there are few or only short-term head-to-head comparisons. Depending on the question, any of these types of studies might provide the best evidence to address gaps in the evidence from head-to-head effectiveness studies. Norris et al. offer further, specific guidance on criteria for including observational studies in CERs in an upcoming article in this series.

2.2.4. Balance of benefits and harms

CERs aim to present benefits and harms for different treatments and tests in a consistent way so that decision-makers can fairly assess the important tradeoffs involved for different treatment or diagnostic strategies. Decision-makers, not the reviewers, must weigh the benefits, harms, and costs of the alternatives. The reviewers, for their part, should seek to present the benefits and harms in a manner that helps with those decisions. The single most important feature of a good CER is that all important outcomes, rather than a selected subset of them, are described.

Expressing benefits in absolute terms (e.g., a treatment prevents one event for every 100 treated patients) rather than in relative terms (e.g., a treatment reduces events by 50%) can also help decision-makers. Reviewers should also highlight where evidence indicates that benefits, harms, and tradeoffs are different for distinct patient groups who, because of their personal characteristics, may be at higher or lower risk of particular adverse effects, or may be more or less susceptible to complications of the underlying condition. Reviews should not attempt to set a standard for how results of research studies should be applied to patients or settings that were not represented in the studies. With or without a CER, these are decisions that must be informed by clinical judgment.

3. Future development of the methods guide

Future articles in this series will look at:

- When and how to use observational studies;
- Assessing the applicability of studies;
- Assessing harms;
- Finding evidence;
- Quantitative synthesis; and
- Rating a body of evidence.

We have also identified several gaps in the methodological literature that will be addressed through new guidance, and future research including methodologies for the assessment of medical tests. Several groups are currently working on developing guidance for medical test assessment that will suggest a framework for the review of medical tests and, among others, will address issues such as when and how to use modeling; how to assess the quality of studies of medical tests; the relevance and consequences of the full range of patient outcomes on decisions to use a medical test; and the assessment of studies of genetic and prognostic tests.

For many of these issues, some variation in practice may persist because of differing opinions about the relative advantages of different approaches and a lack of sufficiently strong empiric evidence to dictate a single method. As further information accumulates, we expect to define more specific requirements related to these issues. We will continue to assess both the ability to implement our recommendations and the validity of the methods that we have adopted—both primary recommendations and secondary concepts introduced in the guidance—as we undertake comparative reviews on a wide assortment of topics. We anticipate that the guidance will continue to evolve as we identify new issues and accumulate experience with new topic areas.

References

