

***Methods Guide for
Comparative Effectiveness Reviews***

Principles in Developing and Applying Guidance



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

Comparative Effectiveness Reviews are systematic reviews of existing research on the effectiveness, comparative effectiveness, and harms of different health care interventions. They provide syntheses of relevant evidence to inform real-world health care decisions for patients, providers, and policymakers. Strong methodologic approaches to systematic review improve the transparency, consistency, and scientific rigor of these reports. Through a collaborative effort of the Effective Health Care (EHC) Program, the Agency for Healthcare Research and Quality (AHRQ), the EHC Program Scientific Resource Center, and the AHRQ Evidence-based Practice Centers have developed a *Methods Guide for Comparative Effectiveness Reviews*. This Guide presents issues key to the development of Comparative Effectiveness Reviews and describes recommended approaches for addressing difficult, frequently encountered methodological issues.

The *Methods Guide for Comparative Effectiveness Reviews* is a living document, and will be updated as further empirical evidence develops and our understanding of better methods improves. Comments and suggestions on the *Methods Guide for Comparative Effectiveness Reviews* and the Effective Health Care Program can be made at www.effectivehealthcare.ahrq.gov.

This document was written with support from the Effective Health Care Program at AHRQ. Neither of the authors has a financial interest in any of the products discussed in this document.

Suggested citation: Helfand M, Balshem H. Principles in developing and applying guidance. In: Agency for Healthcare Research and Quality. *Methods Reference Guide for Comparative Effectiveness Reviews* [posted August 2009]. Rockville, MD. Available at: <http://effectivehealthcare.ahrq.gov/healthInfo.cfm?infotype=rr&ProcessID=60>.

Principles in Developing and Applying Guidance

Authors:

Mark Helfand, M.D.^{a,b}

Howard Balshem, M.S.^a

^aOregon Health and Science University Evidence-based Practice Center, Portland, OR

^bPortland VA Medical Center, Portland, OR

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 or the Veterans Health Administration. Therefore, no statement in this report should be construed as an official position of these entities, the U.S. Department of Health and Human Services, or the U.S. Department of Veterans Affairs.

Principles for Developing Guidance

Key Points

To be useful, Comparative Effectiveness Reviews must:

- Approach the evidence from a clinical, patient-centered perspective.
- Fully explore the clinical logic underlying the rationale for a service.
- Cast a broad net with respect to types of evidence, placing high-quality, highly applicable evidence about *effectiveness* at the top of the hierarchy.
- Present benefits and harms for different treatments and tests in a consistent way so that decisionmakers can fairly assess the important tradeoffs involved for different treatment or diagnostic strategies.

CERs are empirically based whenever possible. When empirical evidence is not available or is inadequate, best practices should be defined to reduce variation among reviewers.

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 decisionmakers with accurate, independent, scientifically rigorous information for comparing the effectiveness and safety of alternative clinical options. CERs have become a foundation for decisionmaking in clinical practice and health policy. To play this important role in decisionmaking, CERs must address significant questions that are relevant to patients and clinicians, and they must use valid, objective, and scientifically rigorous methods to identify and synthesize evidence, applying these methods consistently and in an unbiased and transparent manner.

In this chapter, 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 chapters in this guide describe recommended approaches for addressing difficult, frequently encountered methodological issues. The science of systematic reviews is evolving and dynamic. However, excessive variation in methods among systematic reviews gives the appearance of arbitrariness and idiosyncrasy, which undercuts the goals of transparency and scientific impartiality.

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 the inception of the program, the EPCs have been committed to developing methods for identifying and synthesizing evidence that minimize bias. EPCs 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 AHRQ's *Manual for Conducting Systematic Reviews*,¹

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. This approach helps ensure that a review will address important questions that decisionmakers need answered, and it 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 *Methods Guide* is part of a broader system of safeguards to ensure that reviews produced by the EPCs are 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 *Methods Guide* is a collaborative product of the 14 EPCs with oversight from the Scientific Resource Center (SRC). It serves as a resource for the Effective Health Care Program and scientists employed by AHRQ. To prioritize topics for the *Methods Guide*, we:

- Identified challenges in the production of AHRQ evidence reports and variation among EPCs.
- 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 beneficial⁸ or harmful effects of interventions,⁹ using observational studies,¹⁰ assessing diagnostic tests¹¹ 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 a 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 GRADE (Grading of Recommendations Assessment, Development and Evaluation) approaches.
- 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 *Methods 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 papers and reviewed the published guidance from major bodies producing systematic reviews—most importantly, the Cochrane Collaboration Handbook¹⁴ and the Centre for Reviews and Dissemination manual on conducting systematic reviews.^{15,16}

Principles—Developing Guidance

The fundamental principle used in the development of the *Methods Guide* and subsequent guidance has been that workgroups should use empirical, 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 papers, and information published only in abstract form is an example of evidence-based guidance based on empirical research. Many publications on these topics exist,¹⁷⁻¹⁹ and they form a cohesive and consistent body of evidence upon which recommendations can be made.

On the other hand, structural approaches designed to reduce variation in practice and assure consistency across EPCs have also been adopted. Examples are:

- Centralization at the SRC of activities where EPC proficiency and skill vary, such as searching clinical trial registries and the U.S. Food and Drug Administration (FDA) Web site.
- Adoption of strict policies regarding conflicts of interest.
- Introduction of an editorial review process that provides for an independent judgment of the adequacy of an EPC's response to public and peer review comments

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.

Principles—Conducting Comparative Effectiveness Reviews

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:

- Approaching the evidence from a clinical, patient-centered perspective.
- Fully exploring the clinical logic underlying the rationale for a service.
- Casting a broad net with respect to types of evidence, placing a high value on effectiveness and applicability, in addition to internal validity.
- Presenting benefits and harms for different treatments and tests in a consistent way so that decisionmakers 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

decisionmaking.^{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, timely, objective, and scientifically rigorous and to provide for public participation and transparency.

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). Reviews should also take into account the fact that, for many outcomes and decisions, variation in patients’ values and preferences can and should influence decisions.²⁶ Interviews with patients, as well as 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 for the development of key questions to ensure that CERs are patient centered and also meet the needs of other stakeholders.²¹

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 When properly constructed, it 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 health care 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 decisionmakers.

Figure 1 depicts an analytic framework for evaluating studies of a new enteral supplement to heal bedsores. Key questions are associated with the links (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 included would be 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.

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 comparative effectiveness reviews has brought attention to the danger of over-reliance on randomized clinical trials and to 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 from 1997 through 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 RCTs vary in their usefulness in comparative effectiveness reviews.

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. Effectiveness studies use less stringent eligibility criteria, assess health outcomes, and have longer followup periods than most efficacy trials. Roughly speaking, effectiveness studies measure the degree of beneficial effect in “real-world” clinical settings.⁴⁰ The results of effectiveness studies are more applicable to the spectrum of patients who 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.
- 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 function⁴¹ 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.
- 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.
- 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 due to patient selection or inappropriate comparator or dosing of

comparator; may not address all relevant intended effects; may not address all relevant unintended effects; or have 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 and colleagues offer further specific guidance on criteria for including observational studies in CERs in an upcoming chapter in this *Methods Guide*.

Balance of Benefits and Harms

CERs aim to present benefits and harms for different treatments and tests in a consistent way so that decisionmakers can fairly assess the important tradeoffs involved for different treatment or diagnostic strategies. The decisionmakers, 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 (for example, a treatment prevents one event for every 100 treated patients) rather than in relative terms (for example, a treatment reduces events by 50 percent) can also help decisionmakers. Reviewers should 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 comparative effectiveness review, these are decisions that must be informed by clinical judgment.

Future Development of the Methods Guide

Future chapters in this guide will look at:

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

We have identified several gaps in the methodological literature that will be addressed through new guidance. We have also identified future research that is needed, 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 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

empirical 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 the guidance will continue to evolve as we identify new issues and accumulate experience with new topic areas.

References

1. Woolf SH. Manual for conducting systematic reviews. Agency for Health Care Policy and Research: 1996.
2. Agency for Healthcare Research and Quality. Suggesting a Topic for Effective Health Care Research. 2009. Available at: <http://effectivehealthcare.ahrq.gov/documents/TopicFormRevExample.pdf>. Accessed April 27, 2009.
3. Aschengrau A, Seage GR. Essentials of epidemiology in public health. Bartlett and Jones; 2003.
4. Mrkobrada M, Thiessen-Philbrook H, Haynes RB, et al. Need for quality improvement in renal systematic reviews. *Clin J Am Soc Nephrol* 2008 Jul;3(4):1102–14.
5. Shrier I, Boivin JF, Platt RW, et al. The interpretation of systematic reviews with meta-analyses: an objective or subjective process? *BMC Med Inf Decision Making* 2008;8:19.
6. Egger M, Smith GD. Principles of and procedures for systematic reviews [book chapter]. In: Egger M, Smith GD, Altman DG, editors. *Systematic Review in Health Care: Meta-analysis in Context*. 2nd ed. London, England: BMJ Publishing Group; 2001. p. 23–42.
7. Moher D, Soeken K, Sampson M, et al. Assessing the quality of reports of systematic reviews in pediatric complementary and alternative medicine. *BMC Pediatrics* 2002;2(3).
8. Santaguida PL, Helfand M, Raina P. Challenges in systematic reviews that evaluate drug efficacy or effectiveness [review]. *Ann Intern Med* 2005 Jun 21;142(12 Pt 2):1066–72.
9. Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms [review]. *Ann Intern Med* 2005 Jun 21;142(12 Pt 2):1090–9.
10. Norris SL, Atkins D. Challenges in using nonrandomized studies in systematic reviews of treatment interventions [review]. *Ann Intern Med* 2005 Jun 21;142(12 Pt 2):1112–9.
11. Tatsioni A, Zarin DA, Aronson N, et al. Challenges in systematic reviews of diagnostic technologies [review]. *Ann Intern Med* 2005 Jun 21;142(12 pt 2):1048–55.
12. Hartling L, McAlister FA, Rowe BH, et al. Challenges in systematic reviews of therapeutic devices and procedures. *Ann Intern Med* 2005 Jun 21;142(12 Pt 2):1100–11.
13. Helfand M, Morton S, Guallar E, et al. A guide to this supplement. *Ann Intern Med* 2005 Jun 21, 2005;142(12 Pt 2):1033–4.
14. Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions 4.2.6* [updated September 2006]. In: *The Cochrane Library, Issue 4, 2006*. Chichester, UK: John Wiley & Sons, Ltd.
15. National Health Service Centre for Reviews and Dissemination. *Undertaking systematic reviews of research on effectiveness (CRD Report 4, 2nd ed)*. York, UK: NHS Centre for Reviews and Dissemination. University of York; 2001 March. Report No. 4.
16. National Health Service Centre for Reviews and Dissemination. *Review methods and resources*. York, UK: NHS Centre for Reviews and Dissemination, The University of York; 2007 1-26-07].
17. Egger M, Zellweger-Zahner T, Schneider M, et al. Language bias in randomised controlled trials published in English and German. *Lancet* 1997 Aug 2;350(9074):326–9.

18. Moher D, Fortin P, Jadad AR, et al. Completeness of reporting of trials published in languages other than English: implications for conduct and reporting of systematic reviews. *Lancet* 1996 Feb 10;347(8998):363–6.
19. Scherer RW, Dickersin K, Langenberg P. Full publication of results initially presented in abstracts. A meta-analysis. *JAMA* 1994 Jul 13;272(2):158–62.
20. Slutsky J, Atkins D, Chang S, et al. Comparing medical interventions: AHRQ and the effective health-care program [editorial]. *J Clin Epidemiol* 2008 Sep 30.
21. Whitlock EP, Lopez SA, Chang S, et al. Identifying, selecting, and refining topics for comparative effectiveness systematic reviews: AHRQ and the Effective Health Care Program. *J Clin Epidemiol*; to be published.
22. Helfand M. Using evidence reports: progress and challenges in evidence-based decision making *Health Aff* 2005 Jan-Feb;24(1):123–7.
23. Drummond MF, Schwartz JS, Jönsson B, et al. Key principles for the improved conduct of health technology assessments for resource allocation decisions. *Int J Technol Assess Health Care* 2008;24(03):244–58.
24. Black D. POM + EBM = CPD? [editorial]. *J Med Ethics* 2000 Aug;26(4):229–30.
25. Guyatt GH, Montori VM, Devereaux PJ, et al. Patients at the centre: in our practice, and in our use of language [editorial]. *Evidence-Based Med* 2004;9(1):6–7.
26. Guyatt GH, Cook DJ, Haynes B. Evidence based medicine has come a long way [editorial]. *BMJ* 2004 Oct 30;329(7473):990–1.
27. Bravata DM, McDonald KM, Shojania KG, et al. Challenges in systematic reviews: synthesis of topics related to the delivery, organization, and financing of health care. *Ann Intern Med* 2005 June 21, 2005;142(12 Pt 2):1056–65.
28. Harris RP, Helfand M, Woolf SH, et al. Current methods of the US Preventive Services Task Force: a review of the process. *Am J Prev Med* 2001 Apr;20(3 Suppl):21–35.
29. Whitlock EP, Orleans CT, Pender N, et al. Evaluating primary care behavioral counseling interventions: an evidence-based approach [review]. *Am J Prev Med* 2002 May;22(4):267–84.
30. Woolf SH, DiGuseppi CG, Atkins D, et al. Developing evidence-based clinical practice guidelines: lessons learned by the US Preventive Services Task Force [review]. *Ann Rev Public Health* 1996;17:511–38.
31. Mulrow C, Langhorne P, Grimshaw J. Integrating heterogeneous pieces of evidence in systematic reviews. *Ann Intern Med* 1997 Dec 1;127(11):989–95.
32. Bigby M. Challenges to the hierarchy of evidence: does the emperor have no clothes? [article criticism]. *Arch Dermatol* 2001 Mar;137(Mar):345–6.
33. Devereaux PJ, Yusuf S. The evolution of the randomized controlled trial and its role in evidence-based decision making. *J Intern Med* 2003 Aug;254(2):105–13.
34. Shrier I, Boivin J-F, Steele RJ, et al. Should meta-analyses of interventions include observational studies in addition to randomized controlled trials? A critical examination of underlying principles. *Am J Epidemiol* 2007 Aug 21;166(10):1203–9.
35. Walach H, Falkenberg T, Fonnebo V, et al. Circular instead of hierarchical: methodological principles for the evaluation of complex interventions. *BMC Med Res Methodol* 2006;6(29).
36. Tucker JA, Roth DL. Extending the evidence hierarchy to enhance evidence-based practice for substance use disorders. *Addiction* 2006 Jul;101(7):918–32.
37. Atkins D, Fink K, Slutsky J. Better information for better health care: The Evidence-based Practice Center Program and the Agency for Healthcare Research and Quality. *Ann Intern Med* 2005 June 21, 2005;142(12, Pt 2):1035–41.
38. Shekelle PG, Morton SC, Suttorp MJ, et al. Challenges in systematic reviews of complementary and alternative medicine topics. *Ann Intern Med* 2005 June 21, 2005;142(12 Pt 2):1042–7.
39. Pignone M, Saha S, Hoerger T, et al. Challenges in systematic reviews of economic analyses. *Ann Intern Med* 2005 June 21, 2005;142(12 Pt 2):1073–9.
40. Godwin M, Ruhland L, Casson I, et al. Pragmatic controlled clinical trials in primary care: the struggle between external and internal validity. *BMC Med Res Methodol* 2003;3(28).

41. Fullerton DSP, Atherly DS. Formularies, therapeutics, and outcomes: new opportunities. *Med Care* 2004 Apr;42((4 Suppl)):III39–44.
42. Glasgow RE, Magid DJ, Beck A, et al. Practical clinical trials for translating research to practice: design and measurement recommendations. *Med Care* 2005 Jun;43(6):551–7.
43. Kotaska A. Inappropriate use of randomised trials to evaluate complex phenomena: case study of vaginal breech delivery [review]. *BMJ* 2004 Oct 30;329(7473):1039–42.
44. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: increasing the value of clinical research for decision making in clinical and health policy. *JAMA* 2003 Sep 24;290(12):1624–32.
45. Medical Research Council. A framework for development and evaluation of RCTs for complex interventions to improve health. London, England: Medical Research Council; 2000
46. McAlister FA, Straus SE, Sackett DL. Why we need large, simple studies of the clinical examination: the problem and a proposed solution. CARE-COAD1 group. Clinical Assessment of the Reliability of the Examination-Chronic Obstructive Airways Disease Group. *Lancet* 1999 Nov 13;354(9191):1721–4.
47. Mosteller F. The promise of risk-based allocation trials in assessing new treatments [editorial]. *Am J Public Health* 1996 May;86(5):622–3.

Figure 1. Analytic framework for a new enteral supplement to heal bedsores

