

# Challenges in Systematic Reviews That Evaluate Drug Efficacy or Effectiveness

P. Lina Santaguida, PhD; Mark Helfand, MD, MPH; and Parminder Raina, PhD

Increasingly, consumers, clinicians, regulatory bodies, and insurers are using systematic reviews of drug interventions to select treatments and set policies. Although a systematic review cannot provide all the information a clinician needs to make an informed choice for therapy, it can help decision makers distinguish what claims about effectiveness are based on evidence, identify critical information gaps, describe features of the evidence that limit applicability in practice, and address whether drug effectiveness differs for particular subgroups of patients. To improve the relevance and validity of reviews of drug therapies, reviewers need to delineate clinically important subgroups, specific aims of therapy,

and most important outcomes. They may need to find unpublished trials, studies other than direct comparator (head-to-head) trials, and additional details of published trials from pharmaceutical manufacturers and regulatory agencies. In this paper, we address ways to formulate questions relevant to specific clinical therapeutic aims; discuss types of studies to include in drug efficacy and effectiveness reviews and how to find them; and describe ways to assess applicability of studies to actual practice.

*Ann Intern Med.* 2005;142:1066-1072.

[www.annals.org](http://www.annals.org)

For author affiliations, see end of text.

Systematic reviews are commonly used to evaluate drug therapies. They may focus on individual drugs, a class of drugs, or drug versus nondrug therapies. Health care professionals, consumers, government agencies, and insurers use systematic reviews to aid in treatment decisions, develop guidelines, and derive preferred drug lists and formularies ([www.aarp.org/health/comparedrugs](http://www.aarp.org/health/comparedrugs); [www.crbestbuydrugs.org](http://www.crbestbuydrugs.org)) (1, 2).

The multiple, high-profile uses for results of systematic reviews of drug therapy have brought critical attention to the methods used to conduct them (3). In this article, we address ways for reviewers to formulate questions that are relevant to specific clinical therapeutic aims; discuss types of studies to include in drug efficacy and effectiveness reviews and how to find them; and describe ways to assess applicability of studies to actual practice. We use experiences from Evidence-based Practice Centers (EPCs) to illustrate approaches to each of these challenges.

## CHALLENGE: FORMULATING RELEVANT QUESTIONS

Reviewers should formulate questions that adequately capture the outcomes (benefits and harms), intended therapeutic aims, relevant clinical subgroups that are most important to clinicians and patients, and the relative comparisons to either placebo or other drugs.

### Identify Important Outcomes

Evidence-based Practice Center researchers often work with expert panels to understand the clinical logic underlying beliefs about the advantages and disadvantages of different drugs. These researchers should explore differences in pharmacologic characteristics among drugs because these differences may underlie experts' beliefs about a drug's potential clinical advantages. For example, thiazolidinediones, which were approved for use in type 2 diabetes on the basis of short-term trials of glycemic control, have anti-inflammatory effects and work by increasing insulin sensitivity rather than stimulating insulin secretion.

Because of these characteristics, some experts believe that these drugs might reduce the long-term risk for microvascular disease and cardiovascular events compared with other treatment options.

Researchers from EPCs should consult patients and read studies of patients' preferences to identify pertinent clinical concerns that even expert health professionals may overlook. For example, in a panel meeting about the selective serotonin-receptor agonists (triptans), one patient with migraines, a Medicaid subscriber, made this observation:

Whichever triptan I get, I'm only going to get 4 of them a month. I have more migraines than that. What I want to know is, which triptan is the most reliable if you only take a half a pill, or even less, without having to take the rest of it a couple hours later? I'd rather get pretty good relief for 8 headaches than complete relief for a short time for 3 or 4 (4).

### Identify Therapeutic Aims of Treatment

Misunderstanding the therapeutic aims is a hazard in evaluating efficacy of drug therapy. Dementia therapies, for

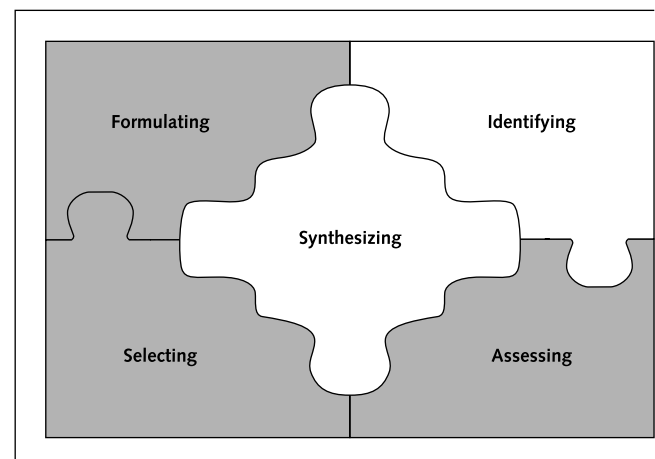
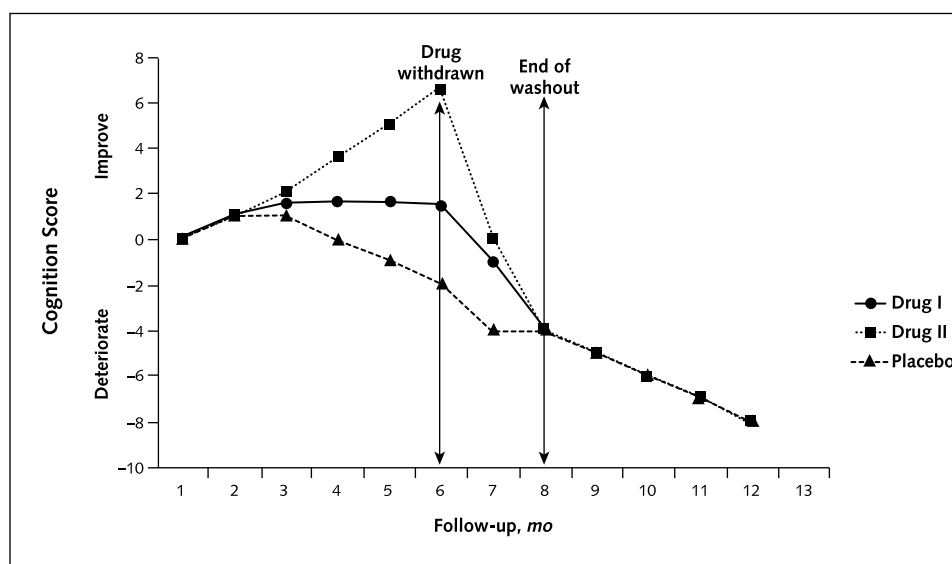


Figure 1. Delay in effects of symptomatic treatment.

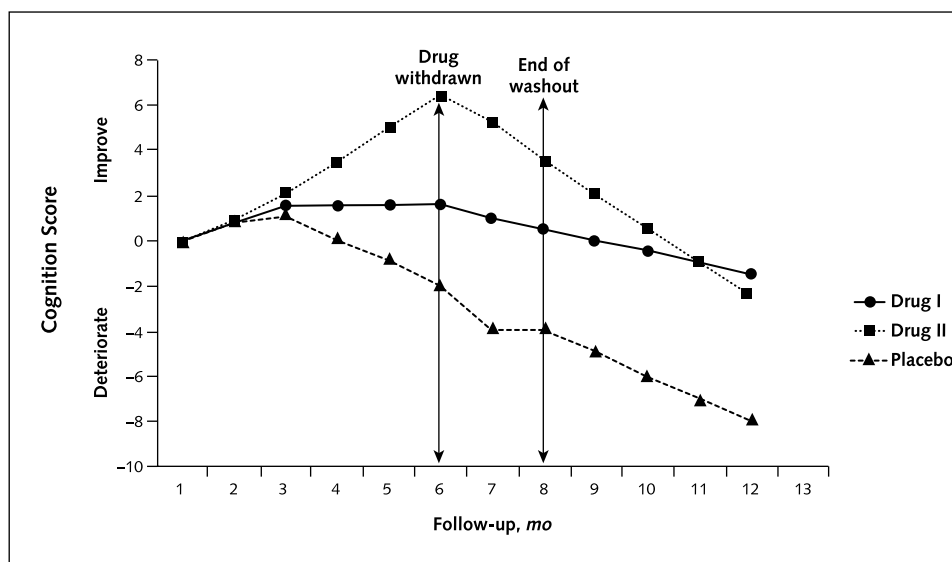


example, are used to alleviate the symptoms of the condition or to modify the underlying disease process. The researcher should identify which of these primary therapeutic goals to address, and then design the eligibility criteria for the systematic review accordingly. A recent systematic review examined the ability of drugs used to treat dementia in order to delay disease onset and prevent the progression of the disease (5). The difference between the 2 therapeutic aims (in the context of dementia) is conceptualized in **Figures 1** and **2**, which show hypothetical responses of patients with dementia to 2 similar drug interventions (I and II) relative to placebo. In these examples, the cognitive abilities of the untreated patients (those who received placebo) with Alzheimer disease decline over time in a manner described by Stern and colleagues (6). For simplicity's sake, the decline in cognition is assumed to be fairly linear; however, the literature has suggested that the rate of decline varies between the different types of dementia and within each of these groups as a function of disease severity (7, 8). In **Figures 1** and **2**, the titration period of 8 weeks (the minimum time required for the drug to be brought to the expected dosage for optimal effect) and the washout period (during which the drug can no longer be acting) have ended at 8 months. **Figure 1** exemplifies the therapeutic aim of symptom relief. Within the active treatment period (first 6 months), the response to drug I depicts the maintenance or stabilization of cognitive function relative to the untreated or placebo; in contrast, the response to drug II suggests improvement (or restoration) of cognition for a short period. Upon withdrawal for patients exposed to either drug I or drug II (following the washout period), the cognition scores declined rapidly to the same rate as that in the untreated (placebo) group; thus, disease pro-

gression was not delayed. In contrast, **Figure 2** shows a delayed rate of decline relative to placebo after the withdrawal of the pharmacologic intervention and, as such, exemplifies disease modification. Comparison of the slopes of the decline of cognition would indicate a greater rate of decline for drug II relative to drug I, but both show delay in progression of the disease effects. Theoretically, the rates of decline in the treatment group will never meet the rate in the untreated group when the pharmacologic agent has truly modified the disease.

Although **Figures 1** and **2** depict idealized responses showing the differences between therapeutic aims, in practice the most effective time interval for causing meaningful change in the outcome (in this example, cognition) and the best time period in which to observe whether the effect is maintained (or lost) are not always known. The difficulty in estimating the ideal time intervals is further compounded when there is uncertainty about the natural rate of disease progression or exactly when the disease can be detected (actual onset of disease vs. classification by a health professional). For many diseases, treatment effects may not be equal across all stages of a disease (mild, moderate, and severe). Moreover, the magnitude of the "improvement" attributed to a drug intervention may depend on the time at which the primary study selected to evaluate the outcome. In some instances, the time selected may reflect a "peak" effect of the drug. For example, drug II in **Figure 2** showed the greatest effect on cognition at 6 months and a lesser effect at 4 months. Thus, researchers critically appraising systematic reviews on drug therapies must judge whether the research question specified the therapeutic aims and must determine whether the time intervals considered were adequate (or justified) for evaluating these intended aims given the specific drugs and diseases.

Figure 2. Delay in effects of disease-progression treatment.



### Specify the Populations and Clinical Subgroups of Interest

Different patients may respond in different ways to the same medication. Some drugs in classes are developed to target certain beneficial effects or avoid certain adverse effects. Regardless, it is unlikely that any single drug in a class will always be the best choice for every patient. In selecting a particular drug for individual patients, several factors, such as the patient's age, sex, race, symptom pattern, other illnesses, other medications, and response to similar medications in the past, may be considered. In formulating key review questions, systematic reviewers should consult experts and recent literature to determine which of these factors to examine.

Increasingly, systematic reviews of drug therapy need to take account of advances in pharmacogenetics and pharmacogenomics. It has been estimated that genetic differences in metabolism, transport, and drug targets account for 20% to 95% of variability in drug effects (9). In some instances, genetic variants have been proven to influence clinical outcomes. A recent dramatic example is the influence of a variant of the  $\alpha$ -adducin gene, which influences renal sodium reabsorption. Among hypertensive patients with the wild-type  $\alpha$ -adducin gene, diuretics were no more effective than other antihypertensive medications. Among patients with the  $\alpha$ -adducin gene variant, risk for myocardial infarction or stroke was 51% lower with diuretics than with other antihypertensive medications (10).

Although there are a few other well-described examples of phenotypic effects of a genetic polymorphism (for example, CYP2D6), most inherited differences in the response to drugs are polygenic and, at present, poorly characterized (11). Clinical studies may reveal variation in drug effects that are not explained by known genetic differences (12). Most commonly, premarketing trials demonstrate ra-

cial differences in elimination half-life, peak concentrations, or other pharmacokinetic features, but the specific genetic differences, and the clinical consequences of these differences, are unclear.

Although a systematic review cannot provide all the information a clinician needs to make an informed choice of therapy, it can address whether drug effectiveness differs for particular subgroups of patients. An EPC review of racial differences in the pharmacologic treatment of heart failure focused on the following most important outcomes of  $\beta$ -blocker and angiotensin-converting enzyme (ACE) inhibitor therapy: reduced mortality and improved quality of life (13). The EPC investigators found that many of the major trials did not include any black patients. Other major trials did not report subgroup results in sufficient detail, so the EPC investigators requested additional analyses or original data from all of the trials that had at least some black participants. They obtained individual-patient data for about half of the trials. The analysis found that, despite evidence from pharmacokinetic studies that black patients respond less than white patients to some ACE inhibitors, no evidence suggested that ACE inhibitor-related mortality reductions in black patients are lesser or greater than those in white patients. For  $\beta$ -blockers, fewer data were available, but overall the mortality reductions for black and white patients were similar: The relative risk reductions were 0.67 (95% CI, 0.39 to 1.16) for black patients versus 0.63 (CI, 0.52 to 0.77) for white patients. These results support the view that EPC reports should consider the clinical implications of pharmacokinetic differences but focus on studies with clinical outcomes.

### Consider Direct and Indirect Comparisons

Ideally, the comparative efficacy or effectiveness of drugs should be evaluated in head-to-head trials (14). Di-

rect comparison studies are not always available, especially for evaluating long-term outcomes such as mortality. In these situations, researchers doing meta-analyses attempt to compare the effectiveness of drugs indirectly, usually from trials that compared one or the other with placebo or a third mode of therapy. Recent examples include comparisons of the effects of different types of antihypertensive drugs on mortality (15) and of triple antiretroviral regimens based on protease inhibitors versus nonnucleoside analogue reverse transcriptase inhibitors on progression to AIDS and death (16). A variety of statistical methods for indirect comparisons are available (17–20). Limited evidence suggests that, when the individual studies are similar and of good quality, and treatment effects are consistent over a variety of comparators, adjusted indirect comparisons usually agree with results of head-to-head comparisons if treatment effects are similar across study samples (21).

## CHALLENGE: FINDING AND SELECTING RELEVANT STUDIES

### Look for Unpublished Data

Identifying unpublished information about a therapy or a class of drugs is essential to assess and correct for publication bias, to obtain information about analyses and results not reported in journal articles, and to assemble a larger population of patients in order to assess the magnitude and significance of effects (22). Publication bias can affect summary estimates of effect in either direction. It occurs when published studies give different results than studies that were rejected from publication, were never submitted, or are otherwise not yet available for review (23). Publication bias does not imply any attempt to mislead. Journals prefer to publish larger, longer, and more positive trials and are unlikely to publish additional trials that confirm the studies that have already been published.

To identify unpublished studies, systematic reviewers should consider soliciting information from manufacturers and regulatory agencies. The U.S. Food and Drug Administration (FDA) sometimes provides information about premarketing pharmaceutical trials (24). The additional information can change estimates of the benefit of a drug or drug class. The Agency for Healthcare Research and Quality (AHRQ) Evidence Report “Treatment of Depression: Newer Pharmacotherapies” (25) did not include unpublished studies (at the time, the information was not available from the FDA). The report concluded that more than 80 studies showed that antidepressants are more efficacious than placebo in treating adults with major depression. Although good evidence indicates that antidepressants are effective, a more recent evaluation of FDA databases for these drugs (26) suggested that fewer than half of antidepressant trials (48% [45 of 93]) were positive; this finding does not correspond to the published literature. It is not clear whether these unpublished studies would have affected the pooled estimates of magnitude of this benefit.

While access to unpublished studies has received the most attention, subtler types of publication bias are also of interest. These include selective reporting of favorable outcomes and, more generally, of favorable types of statistical analyses (sometimes called “publication bias in situ”) (27); duplicate publication of efficacy results from favorable studies; and underreporting of harms. To counter these biases, EPCs should examine all relevant outcomes—benefits and harms—and evaluate studies with a variety of designs and sources (published and unpublished).

### Select Trial Designs That Fit the Therapeutic Aims

Considering types of trial designs (parallel, crossover, or factorial designs) is important in establishing eligibility criteria for the review. An ideal drug development program conducts trials in an ordered sequence: dose tolerance (phase I), dose finding (phase II), dose efficacy (phase III), and postmarketing (phase IV). Because of the pressures on pharmaceutical companies to develop drugs quickly and cost-efficiently, a drug may move into the next phase of development before evidence of the previous phase is completely understood (28). Some systematic reviews of drug therapies may limit studies to a particular phase of the drug development program, typically phase III and higher. Thus, it may be necessary to evaluate earlier drug development trials in order to assess whether later phases of drug study are indeed using appropriate doses in optimum time intervals.

The trial design should match the intended therapeutic aims identified in the research questions. Most premarketing studies use a parallel-group design, meaning that all patients are randomly assigned to alternative intervention groups and are followed for a similar period. This design is best suited to questions about symptomatic treatment and long-term health outcomes, such as mortality and quality of life. Other, less commonly used designs are better for evaluating a drug’s ability to modify the course of a disease. Effects that alter disease course can be distinguished by use of a “withdrawal maneuver” or a “staggered start maneuver” (29, 30). In the former, the drug is simply withdrawn (blinded) at a point at which the separation between the treatment and placebo group is well demarcated. In the latter design, the drug is offered to the patients in the placebo group at the end of the trial.

### Consider Including Observational Studies

Overreliance on randomized trials, especially premarketing trials that assess whether a drug works among highly selected patients, is a frequent and sustained criticism of evidence-based medicine.

However, well-designed observational studies have 2 roles in systematic reviews of drug effectiveness. The first, and most straightforward, is to examine outcomes that are ignored or poorly evaluated in efficacy trials, either because the trials are narrow in scope or because they have a short follow-up period. Nearly all randomized trials of triptans, for example, focus on outcomes related to a few isolated

episodes of migraine. Long-term improvements in work attendance and performance have been examined only in uncontrolled, observational time series studies (4). The other essential role of observational studies is to elucidate and quantify the degree of bias in the efficacy studies. When available, actual evidence about the effect of these biases is powerful. For example, in efficacy trials of interferon- $\alpha$  plus ribavirin or pegylated interferon for hepatitis C, the average rate of sustained viral response ranges from 30% to 40% (31). To see whether these results could be achieved in actual practice, investigators at a large metropolitan county hospital reviewed the charts of 327 patients who had been referred to a liver clinic for further evaluation of hepatitis C (32). Of the 327 patients, 34 had no detectable hepatitis C virus RNA. Of the remaining 293 patients, 210 were not treated, most often because of non-adherence to evaluation procedures or medical or psychiatric contraindications. Of the 83 treated patients, only 13% had a sustained viral response.

Deciding when and how to use observational studies to compare the effectiveness of treatments is nevertheless difficult. First, randomized, controlled trials can incorporate features that make their results widely applicable to everyday practice. “Practical” clinical trials and other effectiveness trials, especially when performed in the setting of a network of community practitioners, can provide better evidence about an intervention’s benefits and risks in everyday practice (33). With the rapid rise in the development of research networks composed of community-based clinicians, such trials are more feasible than ever, and no one disputes that such trials would provide the strongest possible evidence about the comparative effectiveness of therapies. Second, clear or valid criteria for judging the validity of observational effectiveness studies of pharmaceuticals are lacking. Although there may be good agreement in the results of good-quality traditional cohort or case-control studies and large controlled trials, these comparisons did not include observational studies based on large pharmacoepidemiologic databases (33–36).

### CHALLENGE: ASSESSING APPLICABILITY

Uncertainty about the applicability of study samples, settings, and interventions in premarketing trials to practice is a major concern in systematic reviews of drug therapies. Conventional understanding would suggest that “efficacy” studies assess whether a drug works among highly selected patients in closely controlled circumstances. In contrast, “effectiveness” studies aim to recruit (or observe) patients who are likely to be offered the drug, examine clinical strategies that are more representative or likely to be replicated in practice, and measure longer-term or more relevant outcomes.

In practice, characteristics of efficacy and effectiveness studies overlap, and there is not yet consensus on explicit criteria that distinguish these study approaches. Because of

disagreement about which study features are relevant to “efficacy” and which to “effectiveness,” it is important that reviewers and users agree about what characteristics of studies will be examined to determine how “efficacy” versus “effectiveness” will be established; this, in turn, will affect conclusions about the strength of the evidence and judgments about the applicability of the findings.

In a recent article, Rothwell cataloged more than 30 characteristics of controlled trials that can limit their applicability to practice (37). While Rothwell recommends “stricter requirements for the external validity of RCTs [randomized, controlled trials]” conducted by pharmaceutical companies in the premarketing approval process, many of the characteristics he lists arise from legitimate scientific aims. Strict eligibility criteria, as well as placebo and active-drug run-in periods, are designed to exclude patients who are unlikely to complete the randomized phase of a study because they are less likely than others to benefit from, adhere to, or tolerate the treatment under study. Some studies with strict criteria may exclude persons with comorbid conditions, elderly persons, and minority groups. These restrictions are often applied to meet regulatory requirements and to ensure that research funds are expended on persons who are most likely to complete the study and benefit from inclusion.

Regulatory and approval processes for drug use play an important role in how the efficacy or effectiveness of drugs is determined. In drug development programs, manufacturers of the new drugs design and conduct most clinical trials in a manner that will show adherence to the criteria for approval established by the regulatory body. Although this regulatory process preliminarily conceptualizes efficacy and safety, it does not necessarily encompass all the nuances of benefits and harms that are of interest to clinicians and patients. Thus, some drug trials are designed to meet regulatory criteria rather than inform clinical decision making. As a result, some systematic reviews may by default be limited to a subset of outcomes that reflect the influence of regulatory bodies and may not be the intended choice for evaluating efficacy and effectiveness.

### CONCLUSION AND RECOMMENDATIONS

Methods for systematic reviews of drug therapies are evolving. Reviewers should consider several issues when applying the following steps of a systematic review: formulating questions, searching and selecting relevant evidence, and assessing applicability of evidence. We hope that the recommendations listed in the **Table** help make future reviews about drug efficacy more relevant to clinicians and patients. To fully take advantage of the potential for systematic reviews to improve health outcomes, we also recommend research that assesses the validity of observational studies of drug efficacy and that emphasizes applicability of study results to practice.

**Table. Recommendations for Improving Systematic Reviews of Drug Efficacy and Effectiveness**

Consult experts, patients, and recent articles to identify important outcomes (even if there is a paucity of literature).
Establish the primary therapeutic aims (symptomatic versus disease-modifying) and the time frame necessary to evaluate them.
In the absence of head-to-head trials, be aware of the limitations of indirect comparisons.
Specify populations of interest, including key factors such as age, race, symptom pattern, illnesses, other medications, or past responses that might affect efficacy or effectiveness.
Search multiple databases.
Seek unpublished trials and unpublished data about published trials to minimize publication bias.
Use eligibility criteria that ensure a match between study designs and therapeutic aims.
Select observational studies to examine outcomes not evaluated in controlled trials and to assess whether results in controlled trials and actual practice are similar.
Consider differentiating efficacy and effectiveness studies when assessing applicability of study results.

From McMaster University Evidence-based Practice Centre, Hamilton, Ontario, Canada; and Portland Veterans Affairs Medical Center and Oregon Health & Science University Evidence-based Practice Center, Portland, Oregon.

**Grant Support:** This research was performed under contract to the Agency for Healthcare Research and Quality (contract 290-97-0017), Rockville, Maryland. Dr. Raina holds a Canadian Institutes of Health Research (CIHR) Investigator Award and a Premier's Excellence Award (PREA) from the Ontario Provincial Government.

**Potential Financial Conflicts of Interest:** Authors of this paper have received funding for Evidence-based Practice Center reports.

**Requests for Single Reprints:** Parminder Raina, PhD, Department of Clinical Epidemiology & Biostatistics, McMaster University, DTC Room 306, 1280 Main Street West, Hamilton, Ontario L8S 4L8, Canada.

Current author addresses are available at [www.annals.org](http://www.annals.org).

## References

1. Snow V, Weiss K, Wall EM, Mottur-Pilson C. Pharmacologic management of acute attacks of migraine and prevention of migraine headache. *Ann Intern Med.* 2002;137:840-9. [PMID: 12435222]
2. Snow V, Lascher S, Mottur-Pilson C. Pharmacologic treatment of acute major depression and dysthymia. *American College of Physicians-American Society of Internal Medicine.* *Ann Intern Med.* 2000;132:738-42. [PMID: 10787369]
3. Fox DM. Evidence of evidence-based health policy: the politics of systematic reviews in coverage decisions. *Health Aff (Millwood).* 2005;24:114-22. [PMID: 15647221]
4. Helfand M, Peterson K. Drug class review on triptans. Portland, OR: Tertiary Drug Class Review on Triptans; 2004. Accessed at [www.ohsu.edu/drugeffectiveness/reports/documents/Triptans%20Final%20Report%20u2.pdf](http://www.ohsu.edu/drugeffectiveness/reports/documents/Triptans%20Final%20Report%20u2.pdf) on 18 April 2005.
5. Santaguida PS, Raina P, Booker L, Patterson C, Baldassarre F, Cowan D, et al. Pharmacological treatment of dementia. Evidence Report/Technology Assess-

ment No. 97 (Prepared by McMaster University Evidence-based Practice Center under contract 290-02-0020). Rockville, MD: Agency for Healthcare Research and Quality; April 2004. AHRQ publication no. 04-E018-2.

6. Stern RG, Mohs RC, Davidson M, Schmeidler J, Silverman J, Kramer-Ginsberg E, et al. A longitudinal study of Alzheimer's disease: measurement, rate, and predictors of cognitive deterioration. *Am J Psychiatry.* 1994;151:390-6. [PMID: 8109647]
7. Bowler JV, Eliasziw M, Steenhuys R, Munoz DG, Fry R, Merskey H, et al. Comparative evolution of Alzheimer disease, vascular dementia, and mixed dementia. *Arch Neurol.* 1997;54:697-703. [PMID: 9193204]
8. Leber P. Slowing the progression of Alzheimer disease: methodologic issues. *Alzheimer Dis Assoc Disord.* 1997;11 Suppl 5:S10-21; discussion S37-9. [PMID: 9348423]
9. Kalow W, Tang BK, Endrenyi L. Hypothesis: comparisons of inter- and intra-individual variations can substitute for twin studies in drug research. *Pharmacogenetics.* 1998;8:283-9. [PMID: 9731714]
10. Psaty BM, Smith NL, Heckbert SR, Vos HL, Lemaitre RN, Reiner AP, et al. Diuretic therapy, the alpha-adducin gene variant, and the risk of myocardial infarction or stroke in persons with treated hypertension. *JAMA.* 2002;287:1680-9. [PMID: 11926892]
11. Weinshilboum R. Inheritance and drug response. In: Guttmacher AE, Collins FS, Drazen JM, eds. *Genomic Medicine.* Baltimore: Johns Hopkins Univ Pr and *The New England Journal of Medicine;* 2003:41-53.
12. Frackiewicz EJ, Shiovitz TM, Jhee SS. Ethnicity in Drug Development and Therapeutics. London: Greenwich Medical Media; 2002.
13. Shekelle P, Rich M, Morton S, Atkinson S, Maglione M, Heidenreich P, et al. Pharmacologic management of heart failure and left ventricular systolic dysfunction: effect in female, black, and diabetic patients, and cost-effectiveness. Evidence Report/Technology Assessment No. 82 (Prepared by the Southern California-RAND Evidence-based Practice Center under contract 290-97-0001). Rockville, MD: Agency for Healthcare Research and Quality; July 2003. AHRQ publication no. 03-E045.
14. Baker SG, Kramer BS. The transitive fallacy for randomized trials: if A bests B and B bests C in separate trials, is A better than C? *BMC Med Res Methodol.* 2002;2:13. [PMID: 12429069]
15. Psaty BM, Lumley T, Furberg CD, Schellenbaum G, Pahor M, Alderman MH, et al. Health outcomes associated with various antihypertensive therapies used as first-line agents: a network meta-analysis. *JAMA.* 2003;289:2534-44. [PMID: 12759325]
16. Yazdanpanah Y, Sissoko D, Egger M, Mouton Y, Zwahlen M, Chêne G. Clinical efficacy of antiretroviral combination therapy based on protease inhibitors or non-nucleoside analogue reverse transcriptase inhibitors: indirect comparison of controlled trials. *BMJ.* 2004;328:249. [PMID: 14742351]
17. Eddy DM, Hasselblad V, Shachter R. An introduction to a Bayesian method for meta-analysis: the confidence profile method. *Med Decis Making.* 1990;10:15-23. [PMID: 2182960]
18. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med.* 2002;21:2313-24. [PMID: 12210616]
19. Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol.* 1997;50:683-91. [PMID: 9250266]
20. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med.* 2004;23:3105-24. [PMID: 15449338]
21. Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ.* 2003;326:472. [PMID: 12609941]
22. The Cochrane Collaboration. *The Cochrane Reviewers Handbook.* Version 4.1.5 Updated December 2003. The Cochrane Collaboration; 2003.
23. Song F, et al. Publication and related biases. Health Technology Assessment NHS R&D HTA programme. York, United Kingdom: York Publishing Services; 2000;4(10).
24. Chou R, Helfand M. Challenges in systematic reviews that assess treatment harms. *Ann Intern Med.* 2005;142:1090-9.
25. Mulrow CD, Williams JW, Trivedi M, Chiquette E, Aguilar C, Cornell JE. Treatment of depression: newer pharmacotherapies. Evidence Report/Technology Assessment No. 7 (Prepared by San Antonio Evidence-based Practice Center based at the University of Texas Health Science Center at San Antonio under contract 290-97-0012). Rockville, MD: Agency for Health Care Policy and Research; February 1999. AHCPR publication no. 99-E014.
26. Kahn A. Are placebo controls necessary to test new antidepressants and

- anxiolytics? *Neuropsychopharmacology*. 2002;5:193-7.
27. **Phillips CV**. Publication bias in situ. *BMC Med Res Methodol*. 2004;4:20. [PMID: 15296515]
28. **Huster WJ, Enas GG**. A framework establishing clear decision criteria for the assessment of drug efficacy. *Stat Med*. 1998;17:1829-38. [PMID: 9749450]
29. **McDermott MP, Hall WJ, Oakes D, Eberly S**. Design and analysis of two-period studies of potentially disease-modifying treatments. *Control Clin Trials*. 2002;23:635-49. [PMID: 12505242]
30. **Whitehouse PJ, Kittner B, Roessner M, Rossor M, Sano M, Thal L, et al**. Clinical trial designs for demonstrating disease-course-altering effects in dementia. *Alzheimer Dis Assoc Disord*. 1998;12:281-94. [PMID: 9876956]
31. **Chou R, Clark EC, Helfand M**. Screening for hepatitis C virus infection: a review of the evidence for the U.S. Preventive Services Task Force. *Ann Intern Med*. 2004;140:465-79. [PMID: 15023713]
32. **Falck-Ytter Y, Kale H, Mullen KD, Sarbah A, Sorescu L, McCullough AJ**. Surprisingly small effect of antiviral treatment in patients with hepatitis C. *Ann Intern Med*. 2002;136:288-92. [PMID: 11848726]
33. **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;290:1624-32. [PMID: 14506122]
34. **Benson K, Hartz AJ**. A comparison of observational studies and randomized, controlled trials. *N Engl J Med*. 2000;342:1878-86. [PMID: 10861324]
35. **Concato J, Shah N, Horwitz RI**. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med*. 2000;342:1887-92. [PMID: 10861325]
36. **Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al**. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA*. 2001;286:821-30. [PMID: 11497536]
37. **Rothwell PM**. External validity of randomised controlled trials: "to whom do the results of this trial apply?" *Lancet*. 2005;365:82-93. [PMID: 15639683]