Priorities Among Effective Clinical Preventive Services

Methods

Michael V. Maciosek, PhD, Nichol M. Edwards, MS, Ashley B. Coffield, MPA, Thomas J. Flottemesch, PhD, Winnie W. Nelson, PharmD, Michael J. Goodman, PhD, Leif I. Solberg, MD

Abstract: Decision makers want to know which healthcare services matter the most, but there are no well-established, practical methods for providing evidence-based answers to such questions. Led by the National Commission on Prevention Priorities, the authors update the methods for determining the relative health impact and economic value of clinical preventive services. Using new studies, new preventive service recommendations, and improved methods, the authors present a new ranking of clinical preventive services in the companion article. The original ranking and methods were published in this journal in 2001. The current methods report focuses on evidence collection for a priority setting exercise, guidance for which is effectively lacking in the literature. The authors describe their own standards for searching, tracking, and abstracting literature for priority setting. The authors also summarize their methods for making valid comparisons across different services. This report should be useful to those who want to understand additional detail about how the ranking was developed or who want to adapt the methods for their own purposes.

(Am J Prev Med 2006;31(1):90–96)

Introduction

A number of well-known national guidelines outline the clinical preventive services that patients should receive, and are often based on the careful analysis of scientific evidence of effectiveness. For both clinicians and organizational decision makers, however, knowledge that a clinical preventive service is effective is not sufficient to set priorities for increasing the delivery of preventive care. Resources (including clinician and patient time) are limited, and preventive services can differ markedly in their health impact and costs. Clinicians, organizations, and patients need to know which preventive services matter the most. In 2001, a priority ranking of 30 clinical preventive services recommended by the second U.S. Preventive Services Task Force (USPSTF) was presented based on their relative value to the population of the United States.[1] Continuously evolving literature and new recommendations by the current USPSTF make the first ranking increasingly outdated. This article describes the approach used to update that ranking.

From the HealthPartners Research Foundation (Maciosek, Edwards, Flottemesch, Nelson, Goodman, Solberg), Minneapolis, Minnesota; and Partnership for Prevention (Coffield), Washington DC

Address correspondence and reprint requests to: Ashley B. Coffield, MPA, Partnership for Prevention, 1015 18th Street, NW, Suite 200, Washington DC 20036. E-mail: acoffield@prevent.org.

New studies, new recommendations, and improved methods were used to produce an updated ranking. The 2001 methods[2] were adequate for an initial effort to inform priority setting among clinical preventive services, and have been proposed for use in other endeavors.[3,4] However, these new methods take advantage of what was learned previously about data needs and availability, and they address constructive criticisms of the first round, in particular the need for more systematic literature collection and data abstraction.

The National Commission on Prevention Priorities (NCPP), a 30-member panel convened by Partnership for Prevention, and consisting of researchers, health plan executives, employers, and state and federal health officials, guided the study and will guide future updates. The NCPP chose to base the ranking on the same measures used previously, as follows: (1) clinically preventable burden (CPB), which measures a service's health impact, and (2) cost effectiveness (CE), which measures a service's economic value.

The scope of the study chosen by the NCPP applied only to primary and secondary preventive services, including immunizations, screening tests, counseling, and preventive medications offered to asymptomatic people in clinical settings. This included (1) clinical preventive services recommended by the USPSTF through December 2004 for the general asymptomatic population and for persons at high risk of coronary heart disease, and (2) immunizations recommended by

0749-3797/06/$–see front matter
doi:10.1016/j.amepre.2006.03.011

the Advisory Committee on Immunization Practices (ACIP) through December 2004 for the general population.

The primary challenge of priority setting was deriving consistent estimates of a service's CPB and cost effectiveness using disparate data. Obvious differences among immunizations, screening, and counseling complicate this task. Preventive services also differ in the size of their target populations, frequency of delivery, and complexity of achieving the intended health benefits.

A related challenge was evidence collection. The literature provided little methodologic direction about collecting and summarizing the many types of data useful to decision makers. In gathering data for their models, authors of many cost-effectiveness studies conduct reviews and summarize data needed for decision making. However, their search strategies and evidence summaries are rarely systematic or well documented. In the previous study, searches were conducted that were similar to comprehensive cost-effectiveness studies. For the current study, standards were developed to ensure a systematic and transparent process for searching, tracking, and abstracting literature for a priority-setting exercise.

These standards are described here for others who wish to use systematic searches to develop comparable information for decision makers. To provide context to these standards and as a reference for those wishing to understand the ranking, methods used to develop consistent CPB and cost-effectiveness estimates are first summarized; these remain largely unchanged from the earlier analysis.[2] Readers will find a more detailed discussion of these methods in the previous methods report[2] and the complete methods technical report for this update, which is available online.[5]

Estimating a Service's Health Impact and Economic Value

Clinically Preventable Burden

Clinically preventable burden was defined as the total quality-adjusted life years (QALYs) that could be gained in a typical practice if the clinical preventive service were delivered at recommended intervals to a U.S. birth cohort of 4 million individuals over the years of life that a service is recommended. This definition has five embedded principles to promote consistency in the estimation of CPB across clinical preventive services.

1. Clinically preventable burden should include both morbidity and mortality prevented by the service; thus, CPB was measured in terms of QALYs saved. QALYs saved combine years of life gained with improvements in health-related quality of life into a single metric.[6,7] Thus, the number of deaths averted, additional years of life gained per averted death, and the seriousness and duration of illnesses and injuries averted were all considered.
2. Clinically preventable burden should reflect the total potential health benefits from the service among both those currently receiving the service and the rest of the target population. For a service with high effectiveness and high delivery rates, the remaining burden of disease in the U.S. population may be relatively small. By using total health benefits rather than the benefit gained from increasing delivery rates, the overall importance of a service was reflected in the ranking, and effective, well-used services were not undervalued. The estimates reflect the benefit of providing each service, given current delivery rates for any related service in the ranking. Because the childhood primary diphtheria–tetanus–attenuated pertussis (DTaP) series is delivered to 85% of the target population, and 95% receive three doses, the estimates for the tetanus–diphtheria (Td) booster is, essentially, incremental to the provision of the primary series.
3. Clinically preventable burden should take into account expected patient adherence for every service. This was important because it provided a realistic estimate of the expected value of the service when the service is offered as part of usual care. The components of patient adherence included accepting a service once it is offered by a clinician as well as completing follow-up treatments and making needed changes in behavior.
4. Clinically preventable burden was measured for a birth cohort of 4 million that is representative of the U.S. population. The size of the current population for which a service is recommended depends on the size of the birth cohorts that have reached the recommended age range for the service. To reduce this variability among services, CPB was estimated for all services for a hypothetical average birth cohort of 4 million, since recent birth cohorts have been approximately that size.[8] As an alternative, the NCPP considered using a cross-sectional approach to measure CPB in 1 year across the entire age group for whom the service is recommended. This approach would measure the benefit of providing the service to those currently in the recommended age group. The birth-cohort approach was chosen because it reflects the benefit of the service going forward in time, which is consistent with most cost-effectiveness studies.
5. Clinically preventable burden should measure the cumulative benefit of offering the service over the recommended age range at recommended intervals. Some services require only a single intervention (e.g., pneumococcal vaccination) while others require many repetitions (e.g., breast cancer screening) to achieve their full benefit. To account for a service's full benefit, the cumulative benefits of

multiple contacts have been estimated. For example, the CPB of screening for colorectal cancer included the benefits of repeated screenings over several years, and the CPB of tobacco-cessation counseling incorporated the benefits of multiple attempts to engage smokers in cessation activities.

Precision in building CPB estimates that reflect these principles was limited by available data. The previous methods report[2] and the methods technical report[5] provide detail on how consistent CPB estimates were developed while addressing common data limitations. Technical reports for each service that describe the limitations particular to each service in meeting these principles are also available (prevent.org/ncpp). The technical reports for colorectal cancer screening, influenza vaccinations for adults, and tobacco-cessation counseling are summarized in the accompanying articles.[9–11] These services demonstrate the range of challenges in deriving consistent estimates of CPB from available data for screening, immunization, and counseling services.

Cost Effectiveness

The definition of cost effectiveness was the average net cost per QALY gained in a typical practice by offering the clinical preventive service at recommended intervals to a U.S. birth cohort over the recommended age range. Average cost effectiveness is defined as incremental to no provision of the service, but assuming current delivery rates of any related service in the priority ranking. As with CPB, the cost-effectiveness estimates of the Td booster is essentially an incremental estimate to the childhood DTaP series that is widely delivered.

Like CPB, cost effectiveness reflected the provision of the service to the entire target population rather than the marginal cost effectiveness of extending delivery to those not currently receiving the service. Cost effectiveness also incorporates both morbidity and mortality. Costs and QALYs were discounted in the cost-effectiveness ratio (QALYs are not discounted in CPB). These cost-effectiveness estimates also reflected imperfect patient adherence and were estimated over the lifetime of a U.S. birth cohort rather than across the current U.S. cross-section.

The comparability of these cost-effectiveness estimates across services was improved by adhering to the principles of the "reference case" defined by the Panel on Cost Effectiveness in Health and Medicine (PCEHM)[7] and by standardizing all cost-effectiveness ratios to year 2000 dollars. Among the PCEHM reference-case methods that have the largest impact on the cost-effectiveness estimates were the use of a 3% discount rate and the use of a societal perspective, which includes time costs to receive services but not the value of time gained through prevention. The PCEHM reference case excludes the value of time gained (such as productivity gains) through prevented death or illness for most analyses even when using the societal perspective. Although the PCEHM recognized that including the value of such time may be appropriate in some analyses, depending on how quality of life is measured, it was excluded from this current analysis to maintain comparability in the cost-effectiveness estimates among services.

For each service, an estimate of cost effectiveness was developed in one of two ways. Both approaches are demonstrated in at least one of the accompanying articles on colorectal cancer screening, influenza vaccinations, and tobacco-cessation counseling.[9–11] For some services, an existing cost-effectiveness estimate from the literature was used. However, adjustments had to be made to the published cost-effectiveness ratio so that it better reflected the principles outlined above, and thereby was more comparable to cost-effectiveness estimates for other services in the ranking. For example, time costs often had to be added for individuals to receive the service and any follow-up activities. Published cost-effectiveness ratios were adjusted to year 2000 dollars as needed to express all cost-effectiveness ratios in the same base year.

When there was no published cost-effectiveness estimate that yielded an estimate consistent with the principles outlined above, a new cost-effectiveness estimate was produced from a cost-effectiveness model that was based on these CPB estimates. Most of the cost-effectiveness estimates (19 of 25) were based on extensions of the CPB models. The approach to estimating cost-effectiveness ratios differed from standard Markov models in three ways. First, cost-effectiveness ratios were calculated at the population average rather than at the individual level using a hypothetical population with a defined distribution of characteristics and events. Second, the health benefits were calculated as the cumulative expected benefits rather than through year-by-year transition probabilities. For example, the benefits of cholesterol screening were estimated based on average long-term adherence with therapy and the long-term efficacy of therapy in preventing heart disease rather than through yearly probabilities of continued adherence and the projected benefits of 1-year adherence on future heart disease events and associated costs. Third, the estimates were based on the average experience of patients as reported in the effectiveness literature rather than models of each specific clinical pathway that different patients may experience and the probability of each pathway. The accompanying articles on influenza vaccinations and tobacco-cessation counseling demonstrate this approach in a journal article format.[9,10] Readers who wish to see exactly how calculations were performed may consult the technical reports for these and other services.

Table 1. Standardized search strategies: effectiveness and cost effectiveness

Level 1	Level 2	Level 3[a]	Level 4[a]
Search PubMed Limit to English language. Limit to MeSH major terms, title word terms, and phrases. Back to 1992 (01/01/92). Exclude publication types—editorial, comment, news, and letter.	Search PubMed Limit to English language. Limit to text word terms. Back to 1992 (01/01/92). Exclude publication types—editorial, comment, news, and letter.	Search PubMed Limit to English language. Limit to MeSH major terms, title word terms, and text word terms. Back to 1987 (01/01/87). Exclude publication types—editorial, comment, news, and letter.	Search PubMed Limit to English language. Limit to text word terms.
Search Cochrane back to 1992.	References from major articles identified in Level 2.	Other knowledge-based information databases (literature databases).	Search general web
Obtain systematic review articles published back to 1992.			Association websites (American Heart Association, American Cancer Society, etc.)
Obtain articles used as part of the review that were published back to 1987.			
References from major articles identified in Level 1.			Search PubMed for English abstracts from all languages.

[a]Search one or more of the options listed in level.

Overall, the approach incorporated thorough literature reviews, addressed sensitivity analysis, and adhered to all the above mentioned standards consistently, thus making these findings an improvement over the current amalgam of cost-effectiveness reports, which are difficult to compare to one another.

Data Collection and Analysis

Search Strategies

Because thousands of data points must underlie the priority ranking, and resources and time were limited, there were two primary challenges: (1) devise a search strategy that captures the vast majority of evidence while minimizing the resources spent gathering and evaluating estimates that add little precision to the ranking, and (2) provide transparency for the search strategy.

As a matter of practicality in meeting these objectives, two standardized search strategies were developed, one for effectiveness and cost-effectiveness data, and a second for burden of disease and cost data (Tables 1 and 2). Each strategy defined four levels, where Level 1 included the most current literature and data sources, and each subsequent level extended to older sources and those less likely to yield useful data. For each variable, data were searched at Level 1. The quantity and usefulness of data found were evaluated, and the search was extended to the next level if it were likely to improve the estimate of CPB or cost effectiveness.

This approach allowed us to be pragmatic in expending resources on literature searches and at the same time make explicit the point at which this search for evidence was terminated. The search levels used are reported in each service's technical report, and a record was kept (available on request) of the specific search strategies (e.g., Medline keywords and limits) for each service.

Non-U.S. cost-effectiveness studies were not excluded a priori. However, they were not considered until U.S. studies were reviewed and found to be inadequate for use in developing a cost-effectiveness estimate that is comparable to other services. The utility (as defined in the next section) of any available non-U.S. study was weighed against the limitations of inaccurate currency conversion and potentially poor generalizability of resource use from non-U.S. healthcare systems to a priority ranking for U.S. populations. Only one non-U.S. cost-effectiveness study was used in the 2001 ranking and none was used in the current ranking.

Evaluation of Literature

As with other systematic reviews of preventive services,[12–14] literature abstraction forms were used to ensure the consistent evaluation of the effectiveness and cost-effectiveness literature. The abstraction forms were designed to have flexible data-entry capabilities to allow the form to be used across many different services with text boxes for open-ended abstractor comments and few check boxes. The effectiveness abstraction forms included standard elements such as population, environment, statistical significance, and threats to internal and external validity. The forms also prompted reviewers to record data on all components of effectiveness including adherence with offers to receive the

Table 2. Standardized search strategies: burden of disease and cost

Level 1	Level 2	Level 3	Level 4[a]
National data sets	Search PubMed Limit to English language. Limit to MeSH terms and phrases. Back to 1998 (01/01/98). Exclude publication types—editorial, comment, news, and letter.	Search PubMed Limit to English language. Limit to MeSH major terms, title word terms, MeSH terms and phrases. Back to 1990 (01/01/90). Exclude publication types—editorial, comment, news, and letter.	Search PubMed Limit to English language. Limit to text word terms (search only as appropriate i.e., if it would represent current health status).
Government websites (CDC, NIH, AHRQ, etc.)	Association websites (American Heart Association, American Cancer Society, etc.)	Data sources referenced by articles identified in Level 2.	Other knowledge-based information databases (literature databases).
Search PubMed Limit to English language. Limit to MeSH major terms and title word terms. Back to 1998 (01/01/98). Exclude publication types—editorial, comment, news, and letter.	Data sources referenced by articles identified in Level 2.		Search general web.
Data sources referenced by articles identified in Level 1			HealthPartners' data.

[a]Search one or more of the options listed in level.
AHRQ, Agency for Healthcare Research and Quality; CDC, Centers for Disease Control and Prevention; NIH, National Institutes of Health.

service, the portion of cases detected by screening, adherence with follow-up, and the effectiveness of the service or follow-up treatment. When there was insufficient direct evidence on the effectiveness of the preventive services in preventing all important diseases and mortality, searches were expanded to cover these components of effectiveness.

The cost-effectiveness abstraction forms were based on forms developed for the *Guide to Community Preventive Services*,[13] with modifications for use with the literature on clinical preventive services. The sections of the form summarize article information regarding the study population (usually characteristics of a hypothetical population for a model), study or model design as appropriate, the epidemiologic parameters underlying the model (e.g., incidence rates, effectiveness, screening sensitivity, and specificity), the dollar value of costs used in the model, detailed information on what costs were included and how measured (e.g., paid amounts, charges, Medicare payments, cost accounting), model results, and sensitivity analysis.

In pilot work, abstraction forms for data on disease and costs were tested and found to be of too little utility relative to the study resources that their use required. Therefore, the pilot forms were not refined for use in the study. For disease and cost data, the most important pieces of information from articles are about the population in which they were observed, the manner in which costs were measured and disease cases identified, and the components of care included in the cost estimates. This information was entered into a simple spreadsheet by a single reader unless it was part of an abstracted cost-effectiveness article. By comparison, effectiveness and cost-effectiveness articles required a formal abstraction tool and two abstractors to ensure proper recording of the details of intervention design, study design, and analysis.

Both the effectiveness and cost-effectiveness abstraction forms included an evaluation of study usefulness for the purpose of completing a ranking of clinical preventive services. Six criteria were used in the effectiveness abstraction form, and five in the cost-effectiveness abstraction form (Table 3). Reviewers rated each study on a simple 3-point scale for each criterion and entered a brief explanation for the scores. A total "utility score" was calculated as an unweighted average of the individual scores.

Reviewers distinguished between study usefulness and study quality. Usefulness described the value of the article in providing data point(s) that could be used to generate a CPB or cost-effectiveness estimate according to the definitions applied consistently across services.

Table 3. Evaluation of study usefulness

Effectiveness	Cost effectiveness
Study population	**Study population**
Consistency with recommended service age group	Consistency with recommended service age group
Generalizability to U.S. population	Generalizability to U.S. population
Service definition	**Service definition**
Consistency with USPSTF technology	Consistency with USPSTF technology
Consistency with recommended frequency of delivery	Consistency with recommended frequency of delivery
Reporting	**Reporting**
Completeness and clarity	Completeness and clarity
Design	**Design**
Adequacy of study design and study measures to generate reliable estimates	Consistency with PCEHM reference case methods
Adequate sample size	
Implementation	**Sensitivity analysis**
Consistent with design	Completeness and clarity of sensitivity analysis in identifying variables influential to average CE ratio
Evaluation/analysis	
Appropriate statistical analysis	
Adequate control of potential selection bias and other population differences	

CE, cost effectiveness; PCEHM, Panel on Cost Effectiveness in Health and Medicine; USPSTF, U.S. Preventive Services Task Force.

Some markers of study quality were important for this determination, including those related to the internal and external validity of the study.[15–19] However, higher-quality studies may have been less useful for these purposes, and lower-quality studies may have been useful if they provided a key piece of data that was not greatly affected by the study's quality shortcomings.

The utility scores were not used directly to calculate CPB, cost effectiveness, or the ranking. Instead they were used as a "red flag" when reviewing the data that were abstracted from each study. Lower scores warned the study team to carefully review the study before using the study results to estimate CPB or cost effectiveness. In addition to the criteria rating shown in Table 3, studies could be flagged with a "fatal flaw," indicating that the reviewer believed that the study results were not suitable for estimating CPB or cost effectiveness due to a critical quality concern.

Two reviewers abstracted each article. Seven individuals served as abstractors, including three PhDs whose specialty areas include economics and statistics, an MD, a PharmD, and four individuals holding master's degrees in health services research and public health. An adjudication meeting was convened at which reviewers discussed discordant entries and reached agreement on a final adjudicated abstraction form, which included a simple average of the reviewers' utility scores. Agreements were reached based on discussion among the reviewers and a third member of the study team who convened the adjudication meeting. Discordant entries usually fell into one of the following areas: error in data entry, error in interpretation of text or tables, or uncertainty of interpretation due to unclear reporting in the article. Discordant entries of the third type were resolved by mutual agreement on the interpretation that was most likely to be correct.

Discussion

These methods differ from the standard approaches used in many systematic literature reviews. In part, this is because many preventive services were evaluated, rather than a single service, not to determine individual effects, but to determine how effective they are relative to others. In addition, a systematic evaluation of cost effectiveness was required. The simplified models demonstrated in the companion articles on colorectal cancer screening, influenza vaccinations, and tobacco-cessation counseling were designed to ensure consistency across a large number of services under the constraint of limited resources.[9–11]

To accommodate the differing characteristics of many services, a system was designed that balanced consistency in principles with flexibility in application. It is impractical to enumerate the differences in application among the services in a single document. Instead, details are available online in technical reports for each service (prevent.org/ncpp).

The rankings may lack reproducibility because of embedded subjective judgments. To produce the best possible estimates, the methods allowed for judgments to accommodate the diversity of data needs and the disparity in data availability among the different services. Judgments were necessary when identifying the most appropriate data within articles, determining when an estimate from a marginally applicable study adds to or detracts from a small body of evidence, and making decisions about secondary outcomes or treatment options that were too insignificant to the value of the service to justify an extensive literature review. The goal was to limit the impact on final results by keeping subjective judgments within the margin of error inherent in the available data. These decisions have been made explicit in each service's technical report.

Detailed complex models are surely better for microlevel decisions, such as what frequency, which target population, and which technology would produce the greatest benefits and least costs for a particular preventive service. These models address the decision makers' need for data syntheses that summarize many data points in meaningful measures of impact and value, produce comparable results, do not overstate the precision of the estimate, and do all of this within a reasonable period of time and at a reasonable cost.

We plan to use increasingly sophisticated modeling techniques over time as we continually update the ranking for the purpose of providing policymakers with more population-specific information. However, to our knowledge, this continues to be the only effort to produce and summarize comparable estimates across a broad range of recommended preventive services. Alternative methods could be used and should be explored alongside these with the goal of finding the models that best meet decision makers' need for information about which clinical preventive services matter most.

We gratefully acknowledge the guidance of the National Commission on Prevention Priorities. This work was supported by the Agency for Healthcare Research and Quality and the Centers for Disease Control and Prevention.

No financial conflict of interest was reported by the authors of this paper.

References

1. Coffield AB, Maciosek MV, McGinnis JM, et al. Priorities among recommended clinical preventive services. Am J Prev Med 2001;21:1–9.
2. Maciosek MV, Coffield AB, McGinnis JM, et al. Methods for priority setting among clinical preventive services. Am J Prev Med 2001;21:10–9.
3. Holtgrave DR. Extending the methodology of the Committee on Clinical Preventive Service Priorities to HIV-prevention community planning. Am J Prev Med 2002;22:209–10.
4. Vogt TM, Aickin M, Ahmed F, Schmidt M. The Prevention Index: using technology to improve quality assessment. Health Serv Res 2004;39:511–30.
5. Maciosek MV, Edwards NM, Solberg LI, et al. Technical report of the National Commission on Prevention Priorities: methods update for priority setting among effective clinical preventive services, 2005. Available at: prevent.org/ncpp.
6. Muennig PA, Gold MR. Using the years-of-healthy-life measure to calculate QALYs. Am J Prev Med 2001;20:35–9.
7. Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-effectiveness in health and medicine. New York: Oxford University Press, 1996.
8. U.S. Census Bureau. Statistical abstract of the United States: 2003 (123rd edition). Washington DC: U.S. Census Bureau, 2003.
9. Solberg LI, Maciosek MV, Edwards NM, Khandchandani HS, Goodman MJ. Repeated tobacco-use screening and intervention in clinical practice: health impact and cost effectiveness. Am J Prev Med 2006;31:62–71.
10. Maciosek MV, Solberg LI, Coffield AB, Edwards NM, Goodman MJ. The health impact and cost effectiveness of influenza vaccination for adults aged 50 to 64 versus those aged 65 and older. Am J Prev Med 2006;31:72–79.
11. Maciosek MV, Solberg LI, Coffield AB, Edwards NM, Goodman MJ. Colorectal cancer screening: health impact and cost effectiveness. Am J Prev Med 2006;31:80–89.
12. Zaza S, Wright-De Aguero LK, Briss PA, et al. Data collection instrument and procedure for systematic reviews in the Guide to Community Preventive Services. Am J Prev Med 2000;18(suppl 1):44–74.
13. Carande-Kulis VG, Maciosek MV, Briss PA, et al. Methods for systematic reviews of economic evaluations for the Guide to Community Preventive Services. Am J Prev Med 2000;18(suppl 1):75–91.
14. Harris RP, Helfand M, Woolf SH, et al. Third U.S. Preventive Services Task Force. Current methods of the U.S. Preventive Services Task Force: a review of the process Am J Prev Med 2001;20(suppl 3):21–35.
15. Guyatt GH, Sackett DL, Cook DJ. User's guides to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid? Evidence-Based Medicine Working Group. JAMA 1993;270:2598–601.
16. Lohr KN, Carey TS. Assessing "best evidence": issues in grading the quality of studies for systematic reviews. Jt Comm J Qual Improv 1999;25:470–9.
17. Slack MK, Draugalis JR. Establishing the internal and external validity of experimental studies. Am J Health Syst Pharm 2001;58:2173–81.
18. 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.
19. Deeks JJ, Dinnes J, D'Amico R, et al. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7:iii–x, 1–173.

The Editors would like to express their sincerest thanks to the group of anonymous peer-reviewers for these papers, some of whom took on the task of reviewing two or more papers. Their wisdom and guidance helped shape the final form of this impressive body of work. Neither we nor the authors could have accomplished this without their tireless efforts.