

**Best Practices for Conducting Economic Evaluations
in Health Care: A Systematic Review of Quality
Assessment Tools**



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

Best Practices for Conducting Economic Evaluations in Health Care: A Systematic Review of Quality Assessment Tools

Prepared for:

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

Contract No. 290-2007-10061-I

Prepared by:

The Johns Hopkins University Evidence-based Practice Center
Baltimore, MD

Investigators:

Damian G. Walker, Ph.D.
Renee F. Wilson, M.S.
Ritu Sharma, B.S.
John Bridges, Ph.D.
Louis Niessen, M.D., Ph.D.
Eric B. Bass, M.D., M.P.H.
Kevin Frick, Ph.D.

This report is based on research conducted by the Johns Hopkins University Evidence-based Practice Center (EPC) under contract to the Agency for Healthcare Research and Quality (AHRQ), Rockville, MD (Contract No. 290-2007-10061-I). The findings and conclusions in this document are those of the authors, who are responsible for its contents; the findings and conclusions do not necessarily represent the views of AHRQ. Therefore, no statement in this report should be construed as an official position of AHRQ or of the U.S. Department of Health and Human Services.

The information in this report is intended to help health care decisionmakers—patients and clinicians, health system leaders, and policymakers, among others—make well-informed decisions and thereby improve the quality of health care services. This report is not intended to be a substitute for the application of clinical judgment. Anyone who makes decisions concerning the provision of clinical care should consider this report in the same way as any medical reference and in conjunction with all other pertinent information, i.e., in the context of available resources and circumstances presented by individual patients.

This report may be used, in whole or in part, as the basis for development of clinical practice guidelines and other quality enhancement tools, or as a basis for reimbursement and coverage policies. AHRQ or U.S. Department of Health and Human Services endorsement of such derivative products may not be stated or implied.

This document is in the public domain and may be used and reprinted without permission except those copyrighted materials noted for which further reproduction is prohibited without the specific permission of copyright holders.

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

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

Suggested citation: Walker DG, Wilson RF, Sharma R, Bridges J, Niessen L, Bass EB, Frick K. Best Practices for Conducting Economic Evaluations in Health Care: A Systematic Review of Quality Assessment Tools. Methods Research Report. (Prepared by Johns Hopkins University Evidence-based Practice Center under contract No. 290-2007-10061-I.) AHRQ Publication No. 12(13)-EHC132-EF. Rockville, MD: Agency for Healthcare Research and Quality. October 2012. www.effectivehealthcare.ahrq.gov/reports/final.cfm.

Preface

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

To improve the scientific rigor of these evidence reports, AHRQ supports empiric research by the EPCs to help understand or improve complex methodologic issues in systematic reviews. These methods research projects are intended to contribute to the research base in and be used to improve the science of systematic reviews. They are not intended to be guidance to the EPC program, although may be considered by EPCs along with other scientific research when determining EPC program methods guidance.

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

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

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

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

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

William Lawrence, M.D., M.S.
Task Order Officer
Center for Outcomes and Evidence
Agency for Healthcare Research and Quality

Acknowledgments

The EPC thanks Brandyn Lau, Dr. Hosne Begum, Todd Noletto, and Beth Barnett for their assistance with the preparation of this report, and Dr. William Lawrence for his valuable insight throughout the project.

Technical Expert Panel

Anirban Basu
Department of Medicine, University of
Chicago
Chicago, IL

Marthe Gold
City University of New York
New York, NY

Miranda Mugford
University of East Angles Norwich
Norwich, UK

Peter Neumann
Tufts Medical Center
Boston, MA

Steven Teutsch
Merck & Co., Inc.
West Point, PA

Peer Reviewers

Paul Barnett
Stanford University
Menlo Park, CA

Steve Gross
University of California, Berkeley
Berkeley, CA

Karen Kuntz
University of Minnesota,
Minneapolis, MN

David Meltzer
University of Chicago
Chicago, IL

Ian Shemilt
University of Cambridge
Cambridge, UK

Best Practices for Conducting Economic Evaluations in Health Care: A Systematic Review of Quality Assessment Tools

Structured Abstract

Objectives: This report describes the strengths and weaknesses of checklists that have been used to evaluate best practices for conducting and reporting on economic evaluations in health care. We defined checklists as any original listing of specific items that the authors recommended be addressed in the conduct or reporting of an economic evaluation. We focused on how checklists define: (1) the criteria for judging that an economic evaluation is of sufficiently high quality to be useful; (2) the importance of different aspects of the evaluation; and (3) the extent to which high quality with respect to one aspect of an evaluation can compensate for lower quality with respect to another aspect of the evaluation.

Methods: A systematic approach was taken to search the literature through January 2012 for articles relevant to economic evaluations, outcomes, and guidelines for the decisionmaker. The following electronic databases were searched using similar search strategies: MEDLINE, EconLit, CINAHL, Embase, and ISI Web of Science. References of relevant reviews were searched for applicable articles. Experts were queried to ensure that all pertinent articles were included.

Results: Ten peer-reviewed journal articles reported on an original checklist for assessing an economic evaluation. The first was published in 1992, and the last was published in 2011. The number of items in the checklists ranged from 11 to 57. One item, perspective, was a criterion in all 10 of the checklists. Eleven other criteria were included in seven to nine of the checklists: description of the target population, choice of alternatives, study question, study design, measurement, valuation, outcome identification, outcome measurement, adjustment for time variation, sensitivity and uncertainties, presentation of results, generalizability, and incremental analysis. Four of the checklists had evidence of excellent test-retest reliability, but none of the checklists had consistent evidence of excellent inter-rater reliability in two or more studies. Only three of the checklists had evidence of excellent criterion validity, based on comparisons between checklists or comparisons with ratings of experts in health economics.

Conclusion: Several well-developed checklists exist for investigators, reviewers, and journal editors to use in efforts to ensure that economic evaluations and eventual systematic reviews of economic evaluations will be more informative and transparent. The choice of an appropriate checklist should be made with the understanding that quality assessment tools will continue to evolve over time and must improve in reliability and validity for all decisionmakers.

Contents

Introduction	1
Methods	3
Literature Search and Study Selection	3
Organization and Tracking of the Literature Search.....	3
Title Review	4
Abstract Review	4
Article Review	4
Data Abstraction	5
Peer Review	5
Results	6
Results of the Literature Search	6
Description of Applicable Checklists	6
Comparison of Checklists	9
Similarities and Differences.....	9
Reliability and Validity of the Checklists for Assessing the Quality of an Economic Evaluation	10
Discussion	16
What Is Being Measured.....	16
Measured Versus Modeled	17
Scoring Schemes.....	17
Reliability and Validity of the Checklists for Assessing the Quality of an Economic Evaluation	18
Systematic Reviews and Critical Appraisal	19
References	21
Figures	
Figure 1. Results of the Search	7
Tables	
Table 1. Number of Citations Through March 2012 That Cited Published Checklists on the Conduct and Reporting of Economic Evaluations in Health Care	12
Table 2. Comparison of the Contents of Published Checklists for the Conduct and Reporting of Economic Evaluations in Health Care	13
Table 3. Summary of the Basic Characteristics of Published Checklists on the Conduct and Reporting of Economic Evaluations in Health Care.....	14

Appendixes

Appendix A. Definition of Terms

Appendix B. Search Strategies

Appendix C. Title Review Form

Appendix D. Abstract Review Form

Appendix E. Article Inclusion/Exclusion Form

Appendix F. Final Article Triage Form

Appendix G. Checklist Details

Introduction

Guidelines exist for the conduct and review of economic evaluations. In the United States, the systematic review of economic outcomes and the inclusion of economic data in systematic reviews have not been standardized as much as is the case for clinical outcomes.¹ The lack of standardization leads to large variation in the quality of economic evaluations and in the use of economic data in systematic reviews.^{2, 3} This suggests a need to establish some standards in the United States.

Systematic reviews play a critical role in determining the comparative effectiveness of medical interventions and are critical for developing clinical practice guidelines (see Appendix A), efficacy-based coverage decisions, and general health policy.⁴⁻⁷ As such, the processes of searching for and summarizing studies illustrating clinical efficacy and effectiveness have been streamlined in the United States and elsewhere.^{8, 9} While much work is being done to evaluate the comparative effectiveness of medical interventions,^{10, 11} the United States has fallen behind international best practices in using economic data in the comparison of medical interventions for the purposes of policymaking.¹²⁻¹⁶ This is despite calls for integrating cost-effectiveness data routinely in the U.S. health care policy process.^{17, 18}

With increased awareness of the importance of evaluating value for money in health care, the number of published economic evaluations has increased in recent years.¹⁹ As more economic analyses are produced, researchers and policymakers need to have methods to synthesize and interpret the results of multiple analyses that address a single issue; systematic review offers a framework for doing this. However, systematic reviews of economic analyses pose special challenges for those who perform reviews and those who use them.²⁰ Traditional techniques of meta-analysis are not appropriate for many economic analyses, which are often syntheses, as in the case of economic modeling studies, and hence should not be combined as one might combine the results of different randomized controlled trials. Instead, systematic reviews of economic modeling studies are most useful for comparing and contrasting how different investigators have chosen to structure their models and estimate key variables. They can also clarify how results differ between studies based on these different assumptions. Identifying sources of variation across studies can help individual decisionmakers determine which studies best apply to their particular settings and can guide future research by identifying areas of uncertainty. Systematic assessment of study quality can help reviewers interpret individual study results. A little over a decade ago, health economists had “not yet developed a formal methodology for reviewing and summing up evidence from individual economic evaluations...or indeed for assessing whether systematic reviews are possible in this context.”²¹ Today, there are as yet no widely validated methodological criteria to be applied to screening economic studies for inclusion in systematic reviews.²²

The difficulty with developing systems to evaluate best practices for conducting economic evaluations is that each economic evaluation faces a potentially unique set of constraints relative to gold standard sets of recommendations. Systems proposed to date have listed criteria for the assessment of economic evaluations, yet have hardly tested comparative rating and weighting of technical criteria. This comparative weighting and the need to determine whether strength in one area can offset a deficiency in another will be critical to the success of the system that is developed. However, it is not clear that systems that will result in a relative ranking of studies that are not perfect can provide anything other than an approximate estimate of the comparative

validity of differing results. Nevertheless, there is unquestionably an urgent need for improving the design, analysis and reporting of economic evaluations in health care.

The aim of this paper is to review the strengths and weaknesses of checklists that have been used to evaluate best practices for conducting economic evaluations in health care (Appendix A). These checklists were identified in a systematic review process. We defined checklists as any original listing of specific items that the authors recommended be addressed in the conduct or reporting of an economic evaluation. We focused our attention on the ways in which the checklists define: (1) the criteria for judging that an economic evaluation is of sufficiently high quality to be useful; (2) the importance of different aspects of the evaluation; and (3) the extent to which high quality with respect to one aspect of an economic evaluation can compensate for lower quality with respect to another aspect of the evaluation.

Methods

To ensure that everyone is reading with the same expectations, we will begin by providing definitions of reliability and validity. Reliability reflects the consistency of the measurement. Validity reflects whether a measure is measuring what we think we are measuring so that the correct conclusion can be drawn.

Literature Search and Study Selection

Searching the literature involved identifying reference sources, formulating a search strategy for each source, and executing and documenting each search. For the searching of electronic databases, we used medical subject heading (MeSH) terms to build our search strategy. To identify articles that were potentially relevant to our aims, we searched for terms relevant to economic evaluations and outcomes, and checklists and decisionmaking. We chose not to limit our search by study design since it was agreed by the research team and our advisors that many study designs may contain the information we were searching for. We also looked for eligible studies by reviewing the references in pertinent reviews and eligible articles, by querying our experts, and by taking advantage of knowledge shared at core team meetings.

Our comprehensive search included electronic searching of the MEDLINE[®] (using PubMed), EconLit, Cumulative Index to Nursing and Allied Health Literature (CINAHL), Embase, and ISI Web of Science. Search strategies specific to each database were designed to enable the team to focus the available resources on articles that were most likely to be relevant. We developed a core strategy for MEDLINE, accessed via PubMed, on the basis of an analysis of the MeSH terms and text words of key articles identified a priori. We searched for articles published between 1991 and January 2012. The PubMed strategy formed the basis for the strategies developed for the other electronic databases (see Appendix B).

Organization and Tracking of the Literature Search

The results of the initial searches (through September 2009) were downloaded into ProCite[®] version 5.0.3 (ISI ResearchSoft, Carlsbad, CA). Duplicate articles retrieved from the multiple databases were removed prior to initiating the review. From ProCite, the articles were uploaded to SRS 4.0 (TrialStat[®] 2003-2007). SRS is a secure, Web-based collaboration and management system designed to speed the review process and introduce better process control and scientific rigor. In February of 2009, the SRS system was transferred to new owners, Mobius Analytics (Ottawa, Canada). Functionality of the system was unchanged. We used this database to store full articles in portable document format (PDF) and to track the search results at the title review, abstract review, article inclusion/exclusion, and data abstraction levels. The updated search (October 2009 through January 2012) was managed using a different systematic review software system: Distiller SR (Evidence Partners, Ottawa, Ontario, Canada). All procedures were the same and forms used in Distiller were identical to those used in the TrialStat software.

Management systems changed due to the fact that at the time of the update TrialStat was no longer being used by this EPC.

Title Review

The study team scanned all the titles retrieved. Two independent reviewers conducted title scans in a parallel fashion. For a title to be eliminated at this level, both reviewers had to indicate that it was ineligible. If the first reviewer marked a title as eligible, it was promoted to the next review level, or if the two reviewers did not agree on the eligibility of an article, it was automatically promoted to the next level (see Appendix C).

The title review phase was designed to capture as many studies as possible that reported on either the impact of economic evaluations on medical decisionmaking or reported on checklists used to evaluate best practices for conducting economic evaluations in health care. All titles that were thought to address the above criteria were promoted to the abstract review phase.

Abstract Review

The abstract review phase was designed to identify articles that applied to the aims reported in this report or to the aims reported in another report about the impact of economic evaluations on medical decisionmaking. An abstract was excluded at this level if it did not apply to one of these aims or for any of the following reasons: does not apply to economics; no original analysis or data; limited case study of a single policy decision; study focuses on a single condition (no decisionmaking component, policymaking component, or quality of the methods section); methods only; or decisionmaking is at the individual clinician level only. Articles written in a language other than English were not excluded but tagged for further evaluation if the abstract had been translated into English and it appeared to apply to one of the aims (Appendix D).

Abstracts were promoted to the article review level if both reviewers agreed that the abstract could apply to one or more of the aims and did not meet any of the exclusion criteria. Differences of opinion were resolved by discussion between the two reviewers.

Article Review

Full articles selected for review during the abstract review phase underwent another independent review by paired investigators to determine whether they should be included in the full data abstraction. At this phase of review, investigators determined which of the aims each article addressed (see Appendix E). Articles could be excluded for the same reasons as in the abstract review level, as well as: no cost-effectiveness analysis component; cost-effectiveness analysis only without any decisionmaking component; study of cost predictors; or costing study. Differences of opinion regarding article eligibility were resolved through consensus adjudication.

The senior investigators then reviewed the remaining pool of articles to confirm eligibility (see Appendix F). Articles were excluded at this level only if two investigators agreed that an article met one of the exclusion criteria listed above for the same reasons noted on the Article Inclusion/Exclusion Form. Potentially eligible articles were stratified by the level of detail presented in the checklist: original checklist with general applicability; original checklist developed for a specific disease, condition, drug, etc.; partial (original) checklist (containing

some elements of economic evaluation); not an original checklist; or adaptation of an original checklist for other purposes.

Our primary purpose was to identify original checklists or new versions of checklists offering a different perspective than checklists already in print. Articles were excluded for further consideration for the following reasons: adapts an original checklist for other purposes; not an original checklist; original checklist developed for a specific disease or condition; or partial original checklist.

After this elimination process, the team ran a comparison of the articles cited in three “gold standard” checklists: Evers (2003),²³ Chiou (2003),²⁴ and Ungar (2003).²⁵ These three articles identified a number of articles not captured in the literature search. These articles were compared with the list of original checklists and each list was evaluated again for eligibility with additional exclusion criteria added: checklist is not specified (it is implied only); published before 1991 (the year Australia became the first country to formally include economic evidence in national health policymaking); not applicable to a general population; checklist is based on approval, not science; methods or description of a checklist development; not a peer-reviewed report, or is a book; checklist is adapted from an original checklist for another purpose; checklist is specific to a single disease or condition; checklist focuses on pharmacy issues only; is a description or validation of a previously published checklist; or not written in English.

Data Abstraction

Once an article was included at this level, reviewers were given the task to identify the background details of each article, and identify each item detailed in the checklist. This information was entered into an Excel spreadsheet for further analysis by the senior investigators. Data abstraction was checked by a senior investigator for accuracy and completeness.

We focused on the specific criteria that were used to rate studies and the degree of flexibility in interpretation that each system allows. We also recorded the year in which the recommendations for a system were made, the source of the recommendations, the specific criteria, whether the system implied fixed levels of relative importance of the different criteria, and any indication of how often and how successfully the system has been used. We looked for empirical evidence supporting the use of specific items or specific systems in rating an economic evaluation, specifically looking for evidence on the reproducibility (e.g., intra-rater, inter-rater, and internal reliability) and content, construct, or criterion validity of the elements in each rating system.

Peer Review

Throughout the project, the core team sought feedback from seven selected external experts with both academic and industry backgrounds in health economics, policy, and cost effectiveness analysis. A draft of the report was reviewed by a group of external peer reviewers as well as representatives of the Agency for Healthcare Research and Quality (AHRQ). In response to the comments from the peer reviewers, we revised the evidence report and prepared a summary of the comments and their disposition for submission to AHRQ.

Results

Results of the Literature Search

The literature search process identified 19,127 citations that were deemed potentially relevant to the aims of our overall project (see Figure 1). The search strategy used in all search engines was modeled on that which we used in MEDLINE[®], with similar search terms (see Appendix B).

In the title review process, we excluded 15,685 citations that clearly did not apply to the aims of our project. In the abstract review process, we excluded 2,756 citations that did not meet one or more of the eligibility criteria (see Methods section for details). At the article inclusion/exclusion phase, we excluded an additional 511 articles that did not meet one or more of the eligibility criteria, leaving 175 articles identified as possible “checklists.” This pool of articles was further reviewed by the senior investigators on the team and 165 were excluded during this process (Figure 1), leaving a total of 10 articles reporting on original guidelines or checklists.

Description of Applicable Checklists

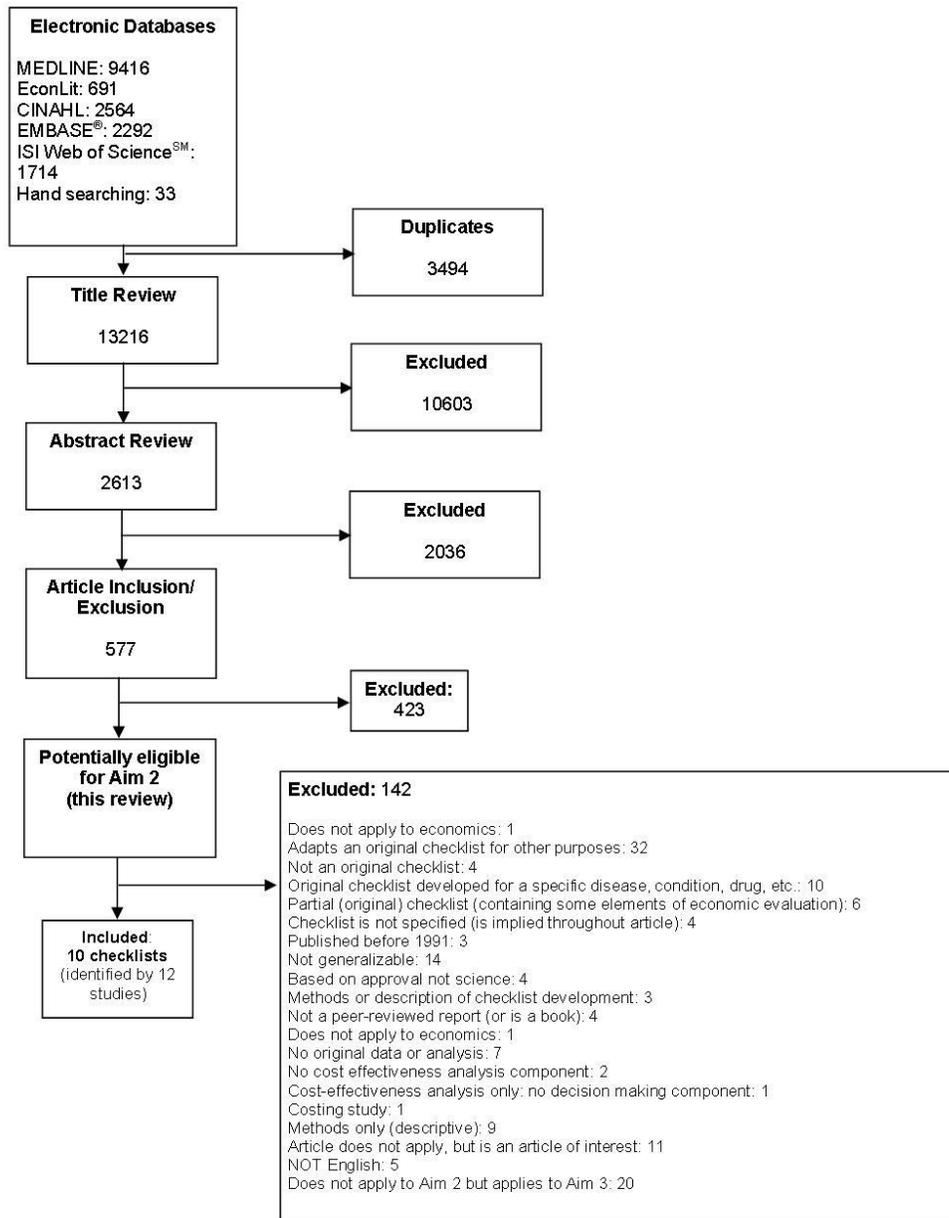
In this section, we describe the eligible articles in chronological order. For each article we provide details on the format of the reported guideline or checklist, its goal, the methods for development, and any validation data.

Adams, 1992.²⁶ To evaluate the completeness of economic analyses in randomized controlled trials, Adams et al. (1992) developed a checklist based substantially on the first edition of the textbook, “Methods for the Economic Evaluation of Health Care Programmes” by Drummond et al., 1987²⁷ (Appendix G, Table 1).

Gerard, 1992.²⁸ The development of these evaluation criteria incorporated a number of views, “all from individuals with experience in the field.” The criteria used by Drummond et al., 1987²⁷ were reference points on the basis of which Gerard revised some and added others. The criteria that evolved from that process were then sent for review by 18 international researchers in the field. Twelve responded with comments and these were taken into consideration. The criteria were tested on a pilot of three cost-utility analyses. The final set of criteria comprised 40 separate judgments to be made per study (Appendix G, Table 2).

Sacristan, 1993.²⁹ Sacristan et al. developed a checklist with 12 sections, each of which includes several subsections, resulting in a total of 40 items. The checklist was devised by analyzing “the most relevant studies on the subject.” After evaluating the corresponding subsections, each section is to be labelled as “correct,” “acceptable,” “doubtful,” “incorrect,” or “not applicable” (Appendix G, Table 3).

Figure 1. Results of the search



CINAHL = Cumulative Index to Nursing and Allied Health Literature

Clemens, 1995.³⁰ In January 1995, the Pharmaceutical Research and Manufacturers of America (PhRMA) adopted a voluntary set of principles to provide guidance for its member companies on the conduct and evaluation of pharmacoeconomic research. The principles were prepared by a working group of pharmacoeconomic scientists from the Task Force on the Economic Evaluation of Pharmaceuticals. The principles were reviewed by a panel of academic experts and outside reviewers at each stage of their development. The PhRMA document consists of a set of broad principles that aim to foster high quality pharmacoeconomic research. While specific recommendations are offered in those methodological areas for which general agreement exists, no attempt was made to force a consensus for those methodological issues which, at the time, were yet to be resolved (and some remain unresolved.) (Appendix G, Table 4).

The U.S. Panel.³¹⁻³³ In the United States, the Panel on Cost-Effectiveness in Health and Medicine³⁴ reviewed the theoretical foundations of cost-effectiveness analysis, current practices, alternative methods, published critiques of cost-effectiveness analyses and criticisms of related general methods and reporting practices. The panel developed recommendations through two and a half years of discussions. Comments on preliminary drafts were solicited from federal government methodologists, health agency officials and academic methodologists (Appendix G, Table 5).

The British Medical Journal (BMJ) Checklist.³⁵ The BMJ set up a working party to develop a quality assessment checklist for use by both referees and authors. Drafts of the checklist were transmitted to health economists and journal editors and were debated at the biannual meeting of the U.K. Health Economists' Study Group in January 1996. The final checklist was based on a broad consensus and contains 35 items under three headings: study design, data collection, and analysis and interpretation of results. This checklist concentrates on full economic evaluations but could also be used for partial economic evaluations, or reports and commentaries on economic evaluations. If items are not applicable to a specific study, a "not appropriate" (NA) response can be stated. The working party admitted that it is not possible to address all the points in an article and that authors can, for example, refer the reader to other published sources (Appendix G, Table 6).

The Pediatrics Quality Appraisal Questionnaire (PQAQ).²⁵ The PQAQ is longer and more detailed than other checklists, and it has been formally validated. It contains 57 items that map into 14 domains. The PQAQ was developed specifically to evaluate the quality of economic appraisals in pediatrics. The authors claim that 9 of the 47 questions are unique to the pediatric population. Scrutiny of these questions suggests this is not the case, however, because it is easy to generalize many of these questions to other populations. Of the 57 questions, 46 items have response options that are scored: 0 if the article fails the criterion or is impossible to judge; 0.5 if the criterion is met partially; or 1 if the criterion is met fully. Ten items refer to descriptive information about the study. The final item is an overall assessment of the quality of the study. This is scored on a 6 point Likert scale, where 1 means excellent and 6 means worthless. A panel of seven experts in health economic evaluation independently assessed potential items for their importance, the clarity of the questions, and the appropriateness of the response categories. Although each of the 46 quantitative items is given a numerical score, the experts involved in the development of the PQAQ cautioned against computing a summary score. Their argument was that each domain was important and a high score on one domain should not be allowed to mask a low score on other domains (Appendix G, Table 7).

The Quality of Health Economic Studies (QHES) List.²⁴ A steering committee comprised of five experts in the field of health economics and three investigators developed a checklist for

economic evaluations from a literature search using Medline, Healthstar, and the Cochrane databases. From existing guidelines and checklists, the committee selected 16 criteria with a “Yes” or “No” format. The selection was made by consensus. Then, weights for each criterion were estimated using a general linear regression (random effects) based on data collected from a conjoint analysis survey of 120 international health economists. The scale was then validated prospectively using a third group of health economists who compared their subjective global assessment of a sample of studies (using a visual analog scale) with scores obtained by the QHES. A study either meets or fails to meet each criterion, thus scoring either the full weighted value or zero for each question. The perfect score for a study is 100 and the lowest score is 0 (Appendix G, Table 8).

The Consensus on Health Economic Criteria (CHEC) List.²³ This is a single checklist, one of the three checklists published by the University of York Centre for Review and Dissemination. An initial item pool divided in 19 categories was first developed by performing a literature search from Medline, Psychlit, Econlit, the Cochrane Library, and the National Health Service Economic Evaluation Databases (NHS EED). The criteria list was then created using the Delphi method. This method made use of a panel of experts on a specific topic to reach a consensus. In a first round, international experts were asked to give their opinion on the categories and the items selected from the literature search. Comments and the resulting list were redistributed among experts until a consensus was reached. Three rounds were sufficient to obtain the final criteria list. More details on the method used can be found in the literature (Evers, 2005²³). The list contains 19 yes or no questions. Authors recommended that if not enough information was available in the article or in other published material to answer a question, a “No” response should be given. It should be noted that this list was not created to analyze the quality of economic evaluations based on modeling studies (Appendix G, Table 9).

A Checklist to Frame Health Technology Assessments for Resource Allocation Decisions. The authors reviewed published literature to identify factors that should be considered when framing health technology assessments for resource allocation decisions (Grutters, 2011³⁶). The checklist contains 11 factors and was finalized in collaboration with clinicians and policy makers (Appendix G, Table 10).

Table 1 shows the number of times the ten checklists have been cited. This provides an indication of the use and importance of each checklist. Although the Drummond and U.S. Panel checklists have been cited most frequently, they were also published over a decade ago, whereas some of the other checklists were published more recently.

Comparison of Checklists

Similarities and Differences

Table 2 compares the checklists using the most recent checklist (the CHEC List)²³ as the reference point. Table 3 details the background characteristics of the checklists. The comparison of instruments showed that they mainly analyze similar items. However, only one item was a criterion in all of the checklists: perspective. In addition, 7 criteria appeared in 11 or more of the checklists: choice of alternatives; perspective; outcome identification; outcome measurement; adjustment for time variation; sensitivity and uncertainties; and incremental analysis. Thus, a strong consensus seemed to exist about some important indicators of quality. On the other hand, four items appeared in five or less of the checklists, suggesting a lack of consensus about their

importance: economic identification, outcome valuation, independence of investigators, and ethics and distribution of effects.

Among the checklists, the number of criteria ranged from 11 in the Grutters et al. Checklist³⁶ to Frame Health Technology Assessments for Resource Allocation Decisions, to 57 in the PQAQ (Ungar and Santos, 2003²⁵). Comparing the tools to the CHEC²³, the PQAQ overlapped the most, with 18 out of the 19 criteria, the exception being “ethics and distribution of effects.” The Grutters checklist³⁶ overlapped the least, matching only 6 out of 19 items, but it had only 11 items in it.

Reliability and Validity of the Checklists for Assessing the Quality of an Economic Evaluation

Gerard (2000)²⁸ assessed the inter-rater reliability of the BMJ economic submissions checklist when reviewing 43 cost-utility analyses in peer-reviewed English language journals in 1996. Proportional agreement between assessors was over 80 percent.

Gerkens (2008)³⁷ compared the BMJ, CHEC, and the QHES instruments. The analysis was based on a review of nine economic evaluations of the surgical treatment of obesity; each paper was assessed independently by two health economists. To compare instruments, the Spearman rank correlation coefficient was calculated for each assessor. Moreover, the test-retest reliability for each instrument was assessed with the intra-class correlation coefficient (ICC). Finally, the inter-rater agreement for each instrument was estimated at two levels: comparison of the total score of each article by the ICC and comparison of results per item by kappa values. The Spearman’s rank correlation coefficient between checklists was usually high ($\rho > 0.70$). Furthermore, test-retest reliability was good for every checklist, that is, 0.98 (95% confidence interval (CI), 0.86 to 0.99) for the BMJ checklist, 0.97 (95% CI, 0.73 to 0.98) for the CHEC list, and 0.95 (95% CI, 0.75 to 0.99) for the QHES instrument. However, inter-rater agreement was poor ($\kappa < 0.40$ for most items and $ICC \leq 0.5$).

Au (2008)³⁸ assessed the reliability of the QHES and PQAQ instruments. Thirty published articles were chosen at random from a recent bibliography of economic evaluations in health promotion. The quality of each of these studies was assessed independently by two raters using each of the two checklists. Inter-rater reliability and the agreement between the checklists were measured using an ICC. Inter-rater reliability was excellent for both checklists (ICC 0.81 for the QHES and 0.80 for the PQAQ). Agreement between the checklists varied (ICC = 0.77 for rater 1 and 0.56 for rater 2).

The developers of the PQAQ instrument²⁵ subjected it to an inter-rater and test-retest reliability assessment. Inter-rater reliability was 0.75 (95% CI, 0.66 to 0.81) and test-retest reliability was 0.92 (95% CI 0.71-0.98).

The developers of the QHES checklist²⁵ sent the checklist to 60 individuals with health economics expertise to validate the tool. These individuals rated the quality of three cost-effectiveness studies on a visual analogue scale, and then evaluated each study using the checklist. They used Spearman rho and Wilcoxon tests to assess convergent validity, and analysis of covariance (ANCOVA) to assess discriminant validity. Convergent validity of the checklist and the general score were shown by the results of the Spearman rho (correlation coefficient = 0.78, $p < 0.0001$) and Wilcoxon test ($p = 0.53$). The latter result’s non-significance was supportive as it suggested a lack of difference in the distribution of the weighted score and

the overall score. Discriminant validity was supported by the ANCOVA ($F(3,146) = 5.97, p = 0.001$).

Table 1. Number of citations through March 2012 that cited published checklists on the conduct and reporting of economic evaluations in health care

	Google Scholar	SCOPUS	Web of Science
Adams ME et al. Economic analysis in randomized control trials. <i>Med Care</i> . 1992; 30: 231-43. ²⁶	151	100	125
Gerard K. Cost-utility in practice: a policymaker's guide to the state of the art. <i>Health Policy</i> . 1992; 21: 249-79. ²⁸	112	68	85
Sacristán JA, Soto J, Galende I. Evaluation of pharmacoeconomic studies: utilization of a checklist. <i>Ann Pharmacother</i> . 1993; 27: 1126-33. ²⁹	70	52	47
Clemens K et al. Methodological and conduct principles for pharmacoeconomic research. <i>Pharmaceutical Research and Manufacturers of America. Pharmacoeconomics</i> . 1995; 8: 169-74. ³⁰	Not listed	52	44
Russell LB et al. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. <i>JAMA</i> . 1996; 276: 1172-7. ³¹	6378	624	600
Siegel JE et al. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. <i>JAMA</i> . 1996; 276: 1339-41. ³³	660	601	548
Weinstein MC et al. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. <i>JAMA</i> . 1996; 276: 1253-8. ³²	1422	1128	1074
Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. <i>BMJ</i> . 1996; 313: 275-83. ³⁵	907	706	626
Ungar WJ, Santos MT. The Pediatric Quality Appraisal Questionnaire: an instrument for evaluation of the pediatric health economics literature. <i>Value Health</i> . 2003; 6: 584-94. ²⁵	13	8	7
Chiou CF et al. Development and validation of a grading system for the quality of cost-effectiveness studies. <i>Med Care</i> . 2003; 41: 32-44. (Quality of Health Economics Studies list) ²⁴	69	55	43
Evers S et al. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. <i>Int J Technol Assess Health Care</i> . 2005; 21: 240-5. ³⁹	100	78	63
Grutters JPC et al. Bridging trial and decision: a checklist to frame health technology assessments for resource allocation decisions. <i>Value in Health</i> . 2011; 14: 777-84. ³⁶	2	1	1

Table 2. Comparison of the contents of published checklists for the conduct and reporting of economic evaluations in health care

Criterion	Adams, 1992 ²⁶	Gerard 1992 ²⁸	Sacristan, 1993 ²⁹	Clemens, 1995 ³⁰	Gold, 1996 ³⁴	Drummond, 1996 ³⁵	Ungar, 2003 ²⁵	Chiou, 2003 ²⁴	Evers, 2005 ²³	Grutters, 2011 ³⁶	SUM
Description of the target population			X	X	X	X	X		X	X	7
Choice of alternatives	X		X	X	X	X	X		X	X	8
Economic study question		X	X	X		X	X	X	X		7
Economic study design	X		X		X	X	X	X	X		7
Follow up period			X	X	X		X	X	X		6
Perspective	X	X	X	X	X	X	X	X	X	X	10
Economic identification	X			X		X	X		X		5
Economic measurement	X	X	X			X	X	X	X		7
Economic valuation	X	X				X	X		X	X	6
Outcome identification	X			X	X	X	X	X	X	X	8
Outcome measurement	X	X	X		X	X	X	X	X		8
Outcome valuation		X				X	X		X		4
Adjustment for time variation	X	X	X	X	X	X	X		X	X	9
Sensitivity and uncertainties	X	X	X	X	X	X	X	X	X		9
Presentation of results		X	X		X	X	X	X	X		7
Generalizability of results		X		X	X	X	X		X		6
Incremental analysis	X		X	X	X	X	X	X	X		8
Independence of investigators							X	X	X		3
Ethics and distribution of effects	X		X		X	X			X		5
Number of the Evers 2005 criteria included in the checklist	12	10	13	11	13	17	18	11	19	6	
Total number of criteria in the checklist	20	37	40	21	37	35	57	16	19	11	

Note: X indicates that the particular list of criteria that is being described in a column includes a criterion related to the issue described in the row.

Table 3. Summary of the basic characteristics of published checklists on the conduct and reporting of economic evaluations in health care

Adams, 1992²⁶	<p>Author Affiliation: The Technology Assessment Group, Department of Health Policy and Management, Harvard School of Public Health, Boston, Massachusetts Year of publication: 1992 Journal: Medical Care Funding: Supported in part by the Alfred P. Sloan Foundation and by grant number HS 05936 from the Agency for Health Care Policy and Research Economic Evaluation: Model-based Target Audience: Investigators using economic analyses in clinical trials Intended use of Checklist: Guidance/Descriptive</p>
Gerard, 1992²⁸	<p>Author Affiliation: University of Aberdeen [Gerard] Year of publication: 1992 Journal: Health Policy Funding: N/A Economic Evaluation: Trial-based and model based CUA Target Audience: Policy makers and researchers who use CUAs (not explicitly stated) Intended use of Checklist: Guidance</p>
Sacristan, 1993²⁹	<p>Author Affiliation: Lilly, S.A., Avda de la Industrial, Madrid Spain Sacristám] Hospital Santa Cruz, Liencres, Cantabria [Soto] Ministry of Health, Madrid, Spain [Galende] Year of publication: 1993 Journal: The Annals of Pharmacotherapy Funding: N/A Economic Evaluation: Model based Target Audience: Researchers, journal editors, and audiences when performing, receiving, reading, or accepting a clinical evaluation study Intended use of Checklist: Guidance</p>
Clemens, 1995³⁰	<p>Author Affiliation: Roche Pharmaceuticals [Clemens] Glaxo, Inc. [Townsend, Oserhaus] Parke-Davis [Luscommbe] Burroughs Wellcome Co [Mauskopf] PhRMA [Babula] Year of publication: 1995 Journal: PharmacoEconomics Funding: PhRMA Task Force on the Economic Evaluation of Pharmaceuticals Economic Evaluation: Model-based Target Audience: Pharmaceutical industry Intended use of Checklist: Guidance</p>
Gold, 1996³⁴	<p>Author Affiliation: Primary affiliations: Department of Maternal and Child Health, Harvard School of Public Health (Siegel), Department of Health Policy and Management, (Weinstein), Institute for Health, Health Care Policy and Aging Research, Rutgers University (Russell), Office of Disease Prevention and Health Promotion, US Public Health Service (Gold). Secondary Affiliation of all authors: US Panel on Cost-effectiveness in Health and Medicine Year of publication: 1996 Journal: JAMA Funding: N/A Economic Evaluation: CEA (Trial- and Model-based) Target Audience: Decision makers who use CEAs (not explicitly stated who) Intended use of Checklist: Guidance</p>

Table 3. Summary of the basic characteristics of published checklists on the conduct and reporting of economic evaluations in health care (continued)

Drummond, 1996³⁵	<p>Author Affiliation: Center for Health Economics, University of York (Drummond) and Ministry of Defence, Army Medical Directorate (Jefferson). The authors wrote the paper on behalf of the BMJ Economic Evaluation Party</p> <p>Year of publication: 1996</p> <p>Journal: British Medical Journal</p> <p>Funding: N/A (BMJ perhaps since it instituted the working party that engaged in this exercise)</p> <p>Economic Evaluation: Trial- and model-based</p> <p>Target Audience: Specialists, non-specialist readers of economic work ; referees and authors; editors</p> <p>Intended use of Checklist: Guidance</p>
Ungar, 2003²⁵	<p>Author Affiliation: Department of Population Health Sciences, The Hospital for Sick Children, Toronto, Ontario, Canada; 2 Department of Health Policy Management and Evaluation, University of Toronto, Toronto, Ontario, Canada</p> <p>Year of publication: 2003</p> <p>Journal: Value in Health</p> <p>Funding: The Canadian Coordinating Office for Health Technology Assessment, from The Hospital for Sick Children Research Institute, and from in-kind support from The Institute of Health Economics</p> <p>Economic Evaluation: Trial- and model-based</p> <p>Target Audience: Decisionmakers regarding allocation of pediatric interventions and services.</p> <p>Intended use of Checklist: Guidance/Descriptive</p>
Chiou, 2003²⁴	<p>Author Affiliation: From Zynx Health Inc., Cedars-Sinai Health System, Beverly Hills, California,</p> <p>Year of publication: 2003</p> <p>Journal: Medical Care</p> <p>Funding: Supported by TAP Pharmaceutical Products, Inc., Lake Forest, Illinois</p> <p>Economic Evaluation: Model-based</p> <p>Target Audience: Clinical decisionmakers and editors/reviewers of medical journals</p> <p>Intended use of Checklist: Guidance</p>
Evers, 2005²³	<p>Author Affiliation: Maastricht University (Evers, Ament and Goossens), Vrije Universiteit University Medical Centre(de Vet and van Tulder)</p> <p>Year of publication: 2005</p> <p>Journal: International Journal of Technology Assessment in Health Care</p> <p>Funding: NS</p> <p>Economic Evaluation: Trial-based</p> <p>Target Audience: Health care professionals, consumers, researchers and policymakers</p> <p>Intended use of Checklist: Guidance</p>
Grutters, 2011³⁶	<p>Author Affiliation: Maastricht University (Gutters) Maastricht University Medical Center (Seferina, Tjan-Heijnen, van Kampen, Joore), Dutch Health Care Insurance Board (Goettsch),</p> <p>Year of publication: 2011</p> <p>Journal: Value in Health</p> <p>Funding: Dutch Organization for Health Research and Development</p> <p>Economic Evaluation: Model-based and Clinical Trials</p> <p>Target Audience: Decisionmakers using health technology assessments for resource allocation</p> <p>Intended use of Checklist: Guidance</p>

CEA: Cost-effectiveness Analysis, CUA: Cost-utility analysis, PhRMA: The Pharmaceutical Research and Manufacturers of America.

Discussion

The first such checklist for economic evaluation of health care was published in 1974,⁴⁰ and some claim that the many efforts at developing guidelines for the conduct and/or reporting of economic evaluations of health care interventions in the 1990s (and 2000s) can be likened to reinventing the wheel.⁴¹ However, this statement ignores that most recent guidelines or checklists do represent useful elaborations of Williams' original checklist that take into account many of the methodological developments that have occurred in the interim. Such adaptations should be expected and welcomed.

When economic evaluations are performed or reported poorly it is difficult to incorporate the economic analyses into an overall assessment of cost effectiveness. By using a checklist consisting of a minimum set of items, such as those described in this review, future economic evaluations could become more transparent, informative, and comparable. Although use of a checklist will not guarantee that the results of an economic analysis are valid, it will help to ensure that an economic analysis has the appropriate components. Thus, checklists should be viewed as being complementary to approaches for maximizing the validity of an analysis, such as performing multiple independent analyses and comparing their results, asking an expert panel to review an economic evaluation, or making an economic model publicly available for review.

To remain relevant for cost effectiveness analysis, quality assessment tools must evolve and continue to improve in reliability and validity for a broad spectrum of decision makers. Just as the tools require refinement as experience with them accumulates, users will also need to address the minimal competencies required of those who use the checklists and who make decisions based upon them. The need for refinement comes from an increasing specification of general criteria as well as an understanding of how general criteria often need to be clarified with sub-questions or are partially met and do not lend themselves to simple yes or no answers. Clearly, more work is needed to help decisionmakers obtain reliable and valid assessments of the quality of economic evaluations so that such information can be incorporated into their decisions in an objective and transparent manner.

What Is Being Measured?

It is now more important than ever to find tools to help filter and interpret enormous amounts of data and thousands of medical literature references. However, checklists can be misapplied if the items do not fit with different types of studies. Inherent flaws in the checklists we reviewed may only be discovered upon repeated use and scrutiny of the results across a wide variety of studies.

Checklists typically cannot separate the quality of reporting from the validity of the design and conduct of a trial. The checklists reviewed in this report have been used primarily as guidance for evaluating the quality of reporting and not as guidance to design studies (although there is no a priori reason why elements of them could not be used for that purpose) or to assess the methodological quality of a study. Many of the checklists contain items that are not directly related to validity but are related to the precision of results (e.g., power calculations) or generalizability (e.g., inclusion and exclusion criteria). When checklist items are weighted and

aggregated into a summary score (see discussion below about this), such limitations can be compounded. Despite the appeal of a summary score to measure quality, research has found that the use of summary scores for clinical trials provides unreliable assessments of validity.⁴² For example, while the objective of the Quality of Health Economic Studies (QHES) tool is to discriminate the quality of studies, many checklist items are more closely related to reporting quality or interpretation of results than internal validity. The checklist places significant weight on issues such as transparency, whether the study objective was clearly stated, and to a lesser extent, the funding source. Important issues related to internal validity were not included in the checklist, such as the nature of randomization and blinding. For example, a cost-effectiveness study that had adequately concealed randomization and was double blinded could receive the same score as a study that inadequately concealed randomization and had no blinding. This is problematic since both poor allocation concealment and blinding have been associated with bias.⁴³ Similarly, the checklist lacks questions to address the internal validity of observational economic evaluations.

It is important to remember that the reliability of a full economic evaluation is in part predicated on its use of reliable effectiveness data. Thus, critical appraisal of an economic evaluation should involve consideration of the sources of bias that may apply to the effectiveness data used in the evaluation.²²

Measured Versus Modeled

In some cases there need to be distinct criteria for evaluating economic data that are generated alongside randomized controlled trials or otherwise measured and data that are the result of modeling. Furthermore, many studies use both methods and those may need a blended set of criteria. Use of the “Drummond checklist”³⁵ and the CHEC list²³ is recommended in Cochrane reviews to inform appraisal of the methodological quality of full economic evaluations conducted alongside single effectiveness studies, and also to inform critical appraisal of partial economic evaluations using the subset of applicable checklist items. If the scope of the critical review of health economics studies encompasses relevant economic modeling studies, then assessments of the methodological quality of such studies will need to be informed by a different checklist. The Drummond checklist and CHEC list are relevant but not sufficient for modelling studies. The “Phillips checklist”⁴⁴ can be used to inform critical appraisal of the methodological quality of economic modeling. Use of this checklist can be supplemented by referring to a published hierarchy of the best available sources to inform each parameter in an economic model.⁴⁵

The CHEC list is designed for clinical trials and observational studies. Consequently, there is no item on model characteristics. Moreover, this checklist does not require that limitations of the studies, particularly observational studies, be specified. On the other hand, the BMJ checklist and the QHES instruments were mainly adapted to modeling studies. Consequently, the item assessing if details of the model were given would not apply to clinical trials. It could be valuable to have a checklist that can be easily adapted to different study designs with specific subquestions for each design.²²

Scoring Schemes

Quantitative measures of quality allow studies to be ranked according to a quality score. One solution is to apply an equal weight for each item, but this strategy does not allow analysts to take into account the relative importance of each criterion. In part for this reason, another instrument was developed: the QHES,²⁴ a grading system in which weightings differ according to the relative importance of each criterion. No empirical evidence has been generated to validate and describe the generalizability of the scoring system. Further, no scoring scheme has used a criterion like “is there a fatal flaw in the study?” to completely eliminate the study from further consideration. Eliminating studies with fatal flaws would make the entire review process more efficient.

Many checklists do not provide a score to enable the simple comparison among studies. Related to this point, some of these checklists and appraisal criteria assume that each criterion shares an equal weight or level of importance. Thus, it is unclear if current instruments have the capability to discriminate between health economic analyses of high and low quality, and whether users of economic literature without specific expertise are able to derive the needed information from the qualitative instruments. Given that many published studies have been viewed as being of poor quality, it may be more discriminating to use the QHES grading system rather than a guideline or checklist with unweighted scoring methods, when evaluating health economic analyses.

These checklists are not limited to simply ascertaining whether the individual items have been addressed or not. Some of the checklists (e.g., Drummond and Jefferson, 1996³⁵) require an assessment of the clarity and appropriateness with which the individual items have been tackled. Others (e.g., Sacristán, 1993²⁹) used a scoring scale for each item with the following possibilities: 4 = correct, 3 = acceptable, 2 = doubtful, 1 = not reported, 0 = incorrect, NA = not applicable. It will often be difficult to choose between a “Yes” or “No” response. Some items regrouped various criteria. Consequently, if only one of the criteria in a group was not addressed, a “no” response should be given, even if the other criteria were addressed adequately. The possibility to use an intermediate value as “partially respected” could thus be interesting. This problem was mostly present with the QHES instrument. For example, one item tested if the time horizon was relevant, if costs and outcomes were discounted, and if the discount rate was justified. It would be interesting to test the impact of subdividing this kind of item.

Reliability and Validity of the Checklists for Assessing the Quality of an Economic Evaluation

Clearly an accepted quality assessment tool could be a substantial contribution if it assists end users of cost-effectiveness studies to discriminate among the exploding body of literature and efficiently identify the studies with superior merit. For producers of such studies, an accepted checklist could establish a clearer standard, potentially encouraging higher quality and greater rigor. To achieve this level of acceptance and use, however, the checklist must have evidence of validity and reliability. Unfortunately, only some of the tools reviewed have been formally validated or shown to be reliable. One difficulty with assessing validity in this case is

that there is no gold standard. Study findings with respect to reliability and inter-instrument comparisons were discussed above.

While Au³⁸ concluded that there is little improvement in reliability to be gained from using more than one rater or more than one assessment of quality, Gerken's³⁷ study illustrated that the results of the quality assessment of economic evaluations were influenced more by the assessor than by the instrument used. These authors therefore concluded that quality assessments should be performed by at least two independent experts with final scoring based on consensus.

Beyond being reliable, the tools must rate studies on how well they actually answer the question posed by the research. The BMJ and QHES tools were assessed for concurrent validity by comparing them with each other (in addition to the Canadian guidelines,⁴⁶ and the Journal of the American Medical Association user's guide⁴⁷). The results of the Au et al. (2008) study suggested that the QHES and the Pediatrics Quality Appraisal Questionnaire (PQAQ) perform equally well. They concluded that the choice between them can therefore be based on other criteria; simplicity and speed of application in the case of one, and detail in the information provided in the case of the other. Both tools as well as the PQAQ and CHEC tools were also assessed against the opinion of experts ("criterion validity") and validated among economists and some decisionmakers. However, we were concerned about potential bias in the methods for selecting these experts, as evidenced by the large representation from the pharmaceutical industry for the QHES tool, and the use of convenience sampling at an annual meeting of a professional society. Hence, the results may overestimate the utility of the instruments.

Acceptance of an instrument as scientifically sound requires that it represent the full content of each of the attributes being measured ("content validity"). While content validity may be relatively easy to assess in established disciplines and with established tests, content validity has proven to be exceedingly difficult to establish with evolving concepts or disciplines, such as cost-effectiveness analysis. The tools reviewed address many of the essential domains by which the soundness of an economic analysis is assessed; however, to the extent that it omits items pertinent to, for example, observational qualitative studies, their content validity might be compromised. Such studies may involve domains that are not captured by the questions in the tools.

The value of an applied instrument is largely determined by its construct validity, a concept more appropriate to a dynamic field such as cost-effectiveness analysis. Construct validity is established over time by the consistency of findings across different users. Such consistency was found by the experts consulted for the QHES study, and, to that extent, the instrument was determined to have adequate construct validity. However, results from its application have yet to be demonstrated across the spectrum of decisionmakers (e.g., health plans, managed care providers, pharmacy benefit managers, hospital pharmacy and therapeutics committees, or researchers), or for the range of the decisions that must be made.

Systematic Reviews and Critical Appraisal

Given that only a few studies are relevant for any given topic, reviewers must set eligibility criteria to exclude low-quality analyses while still retaining enough studies, if possible, to evaluate important differences in model structure and inputs for key variables. Even when a checklist meets high standards, it may not identify studies that meet the need of a specific review. Decisions to include or exclude such studies should therefore be made on the basis of an overall judgment regarding their methodological quality, as well as their relevance in terms of the

economic questions, interventions, populations, and outcomes being studied. This may require an iterative process; in this, analysts may set final selection criteria only after initially assessing the quantity and quality of the extant literature.²²

Research is needed to examine which criteria for assessing the validity of cost-effectiveness studies are important determinants of study results and in what situations. For example, what is the relationship between quality scores (QHES, as an example) and treatment effect (i.e., cost-effectiveness measure)? Do lower scoring studies tend to produce more variable estimates of cost-effectiveness? Do certain components of the checklist (e.g., sufficient time horizon) relate to the size of the treatment effect? Do quality scores vary across study type (i.e., randomized controlled trial, model, and observational study)?

We did not find any reports of empirical research investigating the impact upon the results of a critical review of health economics studies of decisions to include economics studies that meet some but not all standards of methodological quality. However, as with the choice of eligibility criteria relating to quality and design of effectiveness studies, and to the design of health economics studies, it is plausible that use of different data sources for measures of resource use, cost and/or cost-effectiveness has the potential to impact on results.²²

References

1. Pignone M, Saha S, Hoerger T, et al. Challenges in systematic reviews of economic analyses. *Ann Intern Med.* 2005; 142(12 Pt 2):1073-9.
2. Colmenero F, Sullivan SD, Palmer JA, et al. Quality of clinical and economic evidence in dossier formulary submissions. *Am. J. Managed Care.* 2007; 13(7):401-7.
3. Jefferson T, Demicheli V, Vale L. Quality of systematic reviews of economic evaluations in health care. *JAMA.* 2002; 287(21):2809-12.
4. Carande-Kulis VG, Maciosek MV, Briss PA, et al. Methods for systematic reviews of economic evaluations for the Guide to Community Preventive Services. Task Force on Community Preventive Services. *Am J Prev Med* 2000; 18(1 Suppl):75-91.
5. 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; 20(3 Suppl):21-35.
6. Fox DM. Evidence of evidence-based health policy: the politics of systematic reviews in coverage decisions. *Health Aff (Millwood).* 2005; 24(1):114-22.
7. Garber AM. Cost-effectiveness and evidence evaluation as criteria for coverage policy. *Health Aff (Millwood).* 2004;Suppl.Web Exclusives:W4-284-96.
8. Higgins JPT & Green S (editors). *Cochrane Handbook for Systematic Reviews of Interventions.* Version 5.0.0 edition. The Cochrane Collection, 2008.
9. NHS Centre for Reviews and Dissemination. *Undertaking systematic reviews of research on effectiveness: CRD's guidance for carrying out or commissioning reviews (2nd Edition).* York: University of York: NHS Centre for Reviews and Dissemination, 2001.CRD Report 4.
10. Tunis S, Carino TV, Williams RD II, Bach PB. Federal initiatives to support rapid learning about new technologies. *Health Affairs.* 2007; 26(2):w140-9.
11. Tunis SR, Pearson SD. Coverage options for promising technologies: Medicare's 'coverage with evidence development'. *Health Aff (Millwood).* 2006; 25(5):1218-30.
12. American College of Physicians. Information on cost-effectiveness: an essential product of a national comparative effectiveness program. *Ann Intern Med.* 2008; 148(12):956-61.
13. Garber A. A menu without prices. *Ann Intern Med.* 2008; 148(12):964-6.
14. Wilensky G. Cost-effectiveness information: yes, it's important, but keep it separate, please! *Ann Intern Med.* 2008; 148(12):967-8.
15. Berger ML, Teutsch S. Cost-effectiveness analysis: from science to application. *Med Care.* 2005; 43(7 Suppl):49-53.
16. Aspinall S, Good C, Glassman P, Valentino M. The evolving use of cost-effectiveness analysis in formulary management within the Department of Veterans Affairs. *Med Care.* 2005-; 43(7):II, 20-ii-26.
17. Neumann PJ, Palmer JA, Daniels N, Quigley K, Gold MR, Chao S. A strategic plan for integrating cost-effectiveness analysis into the US healthcare system. *Am J Manag Care.* 2008; 14(4):185-8.
18. Eddy D. Reflections on science, judgment, and value in evidence-based decision making: a conversation with David Eddy by Sean R. Tunis. *Health Aff (Millwood).* 2007; 26(4):w500-15.
19. Neumann PJ, Greenberg D, Olchanski NV, et al. Growth and quality of the cost-utility literature, 1976-2001. *Value Health.* 2005; 8(1):3-9.
20. O'Donnell JC, Pham SV, Pashos CL, et al. Health technology assessment in evidence-based health care reimbursement decisions around the world: an overview. *Value Health.* 2009;12 Suppl.2:S1-5.
21. Jefferson T, Mugford M, Gray A, Demicheli V. An exercise on the feasibility of carrying out secondary economic analyses. *Health Econ.* 1996; 5(2):155-65.
22. Tobin GA, Begley CM. Methodological rigour within a qualitative framework. *J Adv Nurs.* 2004; 48(4):388-96.
23. Evers S, Goossens M, de Vet H, et al. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. *Int J Technol Assess Health Care.* 2005; 21(2):240-5.

24. Chiou CF, Hay JW, Wallace JF et al. Development and validation of a grading system for the quality of cost-effectiveness studies. *Med Care*. 2003; 41(1):32-44.
25. Ungar WJ, Santos MT. The Pediatric Quality Appraisal Questionnaire: an instrument for evaluation of the pediatric health economics literature. *Value Health*. 2003; 6(5):584-94.
26. Adams ME, McCall NT, Gray DT, Orza MJ, Chalmers TC. Economic analysis in randomized control trials. *Med Care*. 1992; 30(3):231-43.
27. Drummond MF, Sculpher ML, Torrance GW. et al. *Methods for the Economic Evaluation of Health Care Programmes*, 3rd Ed. Oxford Medical Publications, Oxford University Press, Inc. New York, NY. 2005.
28. Gerard K. Cost-utility in practice: a policy maker's guide to the state of the art. *Health Policy*. 1992; 21(3):249-79.
29. Sacristan JA, Soto J, Galende I. Evaluation of pharmacoeconomic studies: utilization of a checklist. *Ann Pharmacother* .1993; 27(9):1126-33.
30. Clemens K, Townsend R, Luscombe F, Mauskopf J, Osterhaus J, Bobula J. Methodological and conduct principles for pharmacoeconomic research. *Pharmaceutical Research and Manufacturers of America. Pharmacoeconomics* 1995; 8(2):169-74.
31. Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. *JAMA*. 1996; 276(14):1172-7.
32. Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA*. 1996; 276(15):1253-8.
33. Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. *JAMA*. 1996; 276(16):1339-41.
34. Gold MR, Russell LB Siegel JE Weinstein MC. *Cost-effectiveness in health and medicine*. Oxford: Oxford University Press; 1996.
35. Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ*. 1996; 313(7052):275-83.
36. Grutters JP, Seferina SC, Tjan-Heijnen VC, van Kampen RJ, Goettsch WG, Joore MA. Bridging trial and decision: a checklist to frame health technology assessments for resource allocation decisions. *Value Health* 2011; 14(5):777-84.
37. Gerkens S, Crott R, Cleemput I, et al. Comparison of Three Instruments Assessing the Quality of Economic Evaluations: a Practical Exercise on Economic Evaluations of the Surgical Treatment of Obesity. *Int J Technol Assess Health Care*. 2008; 24(3):318-25.
38. Au F, Prahardi S, Shiell A. Reliability of two instruments for critical assessment of economic evaluations. *Value Health*. 2008; 11(3):435-9.
39. Agha Z, Lofgren RP, VanRuiswyk JV. Is antibiotic prophylaxis for bacterial endocarditis cost-effective? *Med Decis Making* 2005; 25(3):308-20.
40. Williams A. The cost-benefit approach. *Br Med Bull*. 1974; 30(3):252-6.
41. Maynard A. Economic evaluation techniques in healthcare. *Reinventing the wheel? Pharmacoeconomics*.1997; 11(2):115-8.
42. Juni P, Witschi A, Bloch R, Egger M. The hazards of scoring the quality of clinical trials for meta-analysis. *JAMA*. 1999; 282(11):1054-60.
43. Lincoln TM, Rief W. How much do sample characteristics affect the effect size? An investigation of studies testing the treatment effects for social phobia. *J Anxiety Disord*. 2004; 18(4):515-29.
44. Philips Z, Bojke L, Sculpher M, et al. Good practice guidelines for decision-analytic modelling in health technology assessment: a review and consolidation of quality assessment. *Pharmacoeconomics*. 2006; 24(4):355-71.
45. Cooper N, Coyle D, Abrams K, Mugford M, Sutton A. Use of evidence in decision models: an appraisal of health technology assessments in the UK since 1997. *J Health Serv Res Policy*. 2005; 10(4):245-50.

46. Baladi J-F, Menon D, Otten N. Use of economic evaluation guidelines: 2 years' experience in Canada. *Health Econ (GBR)*. 1998; 7(3):221-7.

47. Drummond M, Richardson W, O'Brien B, Levine M, Heyland D. Users' guides to the medical literature. XIII. How to use an article on economic analysis of clinical practice. A. Are the results of the study valid? *JAMA*. 1997; 277(19):1552-7.

Appendix A. Definition of Terms

Guidelines give instructions on how to adequately perform or report a study. A researcher performing an economic evaluation may use guidelines in designing and conducting a study.

A checklist is used after the study has been performed and after the results have been published to see if the study met specific guidelines. Of course, there will be a large overlap between guidelines and checklists for the same type of studies.

Criteria lists are used in systematic reviews to consider the methodological quality of the studies included. The criteria usually relate to the internal validity of the studies, but in some criteria lists external validity items are also included. Again, there may be some overlap between criteria lists and guidelines or checklists in a specific research field, but criteria lists are usually more compact and serve a different purpose. That is, the methodological quality assessed by a criteria list should indicate how high or low the chance is that the results of the study are inaccurate or misleading

Appendix B. Search Strategies

Search strategies

Database	Search			Results
PubMed	A	B	C	
	"Cost –benefit analysis"[mh] "Cost of illness"[mh] "economic evaluation"[tiab] "economic outcomes"[tiab] (analysis[tiab] AND (cost[tiab] OR economic[tiab])) "cost effectiveness"[tiab]	"decision making"[mh] "Health Policy"[Mesh] "Decision Making"[tiab] reimbursement[tiab] "Evidence-Based Medicine"[mh] "Evidence-Based Medicine"[tiab] "Technology Assessment, Biomedical"[mh] "Technology Assessment"[tiab] formularies[mh] guideline*[tiab] recommend*[tiab]	Evaluations[tiab] "evaluation"[tiab] "qualitative"[tiab] "focus group"[tiab] Interview[tiab] observation[tiab] outcomes[tiab] analysis[tiab] analyses[tiab]	
	A and B and C (limited 1991 to present)			12219
EconLit and CINAHL	A	B	C	
	TX "cost-benefit analysis" TX "cost of illness" TX "economic evaluation" TX "economic outcomes" TX "cost effectiveness" TX "cost analysis" TX "economic analysis"	TX "Decision making" TX "health policy" TX reimbursement TX "evidence-based medicine" TX "Technology assessment" TX formulary TX guideline	TX evaluation TX qualitative TX "focus group" TX interview TX observation TX outcome TX analysis	
	A and B and C (limited 1991 to present)			5555
EMBASE	A	B	C	
	'cost of illness':ti,ab 'cost benefit analysis':ti,ab 'economic evaluation':ti,ab 'economic outcomes':ti,ab ('analysis':ti,ab AND ('cost':ti,ab OR economic:ti,ab)) 'cost effectiveness':ti,ab AND [humans]/lim	'decision making':ti,ab 'health care policy':ti,ab 'evidence based medicine':ti,ab formulary:ti,ab reimbursement:ti,ab 'technology assessment':ti,ab guideline:ti,ab AND [humans]/lim	evaluation:ti,ab evaluations:ti,ab qualitative:ti,ab 'focus group':ti,ab interview:ti,ab observation:ti,ab outcome:ti,ab analysis:ti,ab AND [humans]/lim	
	A and B and C (limited 1991 to present)			3296
ISI Web of Science	A	B	C	
	TS=("cost benefit analysis" OR "cost of illness" OR "economic evaluation" OR "economic outcome" OR "cost effectiveness")	TS=("decision making" OR "health policy" OR "reimbursement" OR "evidence based medicine" OR "technology assessment" OR "formulary" OR "guideline" OR "recommendation")	TS=(evaluation OR qualitative OR "focus group" OR interview OR observation OR outcomes OR analysis)	

	A and B and C (limited 1991 to present)	3881
	TOTAL	24984
	overlap between databases	5857
	Reviewed total	19127

CINAHL: Cumulative Index to Nursing and Allied Health Literature

Appendix C. Title Review Form

Previewing Only: You cannot submit data from this form



Previewing at Level 1

Refid: 1, Takahashi, T., Saegusa, Y., Takimoto, Y., and Shiragami, M., Pharmacoeconomic analysis of hypertriglyceridemia treatment at the medical institutions, *Yakugaku Zasshi*, 128(12), 2008, p.1783-9
State: Excluded, Level: 1

Submit Data

1. Does this article **POTENTIALLY** apply to Aim 2 or Aim 3?

Aim 2. *Systems* that have been used to evaluate best practices for conducting economic evaluations in health care.

Aim 3. Studies addressing the issues of using *economic outcomes in policy and decision making*.

- Potentially Eligible for Aim 3 (or Aim 2)
- Definitely not eligible for Aim 3 (or Aim 2)
- Unclear

[Clear Selection](#)

Submit Data

Form took 0.203125 seconds to render
Form Creation Date: Not available
Form Last Modified: Jan 7 2009 4:16PM

Appendix D. Abstract Review Form

Previewing Only: You cannot submit data from this form



Previewing at Level 2

Refid: 1, Takahashi, T., Saegusa, Y., Takimoto, Y., and Shiragami, M., Pharmacoeconomic analysis of hypertriglyceridemia treatment at the medical institutions, *Yakugaku Zasshi*, 128(12), 2008, p.1783-9
state: Excluded, Level: 1

SRS Form

<http://www.srsnexus.com/d2d/ul1/review.asp?mode=previewMode&artic..>

..

Keywords:

No keywords available

Abstract:

It has been demonstrated that HMG-CoA reductase inhibitors effectively decrease low density lipoprotein and total cholesterol levels, and presently, HMG-CoA reductase inhibitors are most widely used in hyperlipidemia treatment. On the other hand, it

Does this article POTENTIALLY apply to Aim 2 or Aim 3? Aim 2. *Systems or Guidelines* that have been used to evaluate best practices for conducting economic evaluations in health care.

Aim 3. *Studies addressing the issues of using economic outcomes in policy and decision making.* Include clinical decision making if the decision is not made on the

individual clinician level.

has been demonstrated that fibrate agents decrease triglyceride levels more effectively compared to HMG-CoA reductase inhibitors. A cost-effectiveness study comparing fenofibrate, a fibrate agent, and atorvastatin was therefore conducted in hypertriglyceridemia patients. Referring to an analytical method published in the UK, the percentage of patients received fenofibrate and atorvastatin treatments at each dose level was estimated from prescription records at the medical institutions investigated. Changes in the total cholesterol and triglyceride values after the drug administration were investigated examining published reports. Based on the said data, the treatment effectiveness was measured by the percentage of patients who achieved the target lipid levels. The treatment costs were estimated based on the number of patients investigated and reimbursement prices of the drugs. The incremental cost-effectiveness ratio of fenofibrate in decreasing triglyceride levels was dominant over atorvastatin. The incremental cost-effectiveness ratio of atorvastatin in decreasing low density lipoprotein cholesterol levels was JPY 69911. This provides a model for choosing drug treatments that reflects clinical practices at medical institutions by substituting figures for individual cases.

DO NOT enter answers for more than one question. This will result in an automatic conflicts. Comments are neutral and can be included with the answer to any of the questions on this form.

1.

Yes, this article **potentially** applies to:
both Aim 2 and Aim 3 can be checked if applicable

Aim 2

Aim 3

2.

No, this abstract **does not apply** to Aim 2 and Aim 3 for the following reason(s):

Does not apply to economics

No original analysis or data (this exclusion includes systematic reviews, commentary, or editorials)

Limited case study of a single policy decision: **exclude only if** there is insufficient information about how the economic data was used in decision making process

Study focuses on a single condition: no decision making component, policy making component, or quality of the methods discussion

Cost-effectiveness analysis only: no decision making component

Other (specify)

Methods only (descriptive)

Decision making is at the individual clinician level ONLY
3.

Unclear **or** no abstract available

Unclear. Cannot determine if article applies to Aim 2 or Aim 3. **INCLUDE** (move to next level)

In a language other than English
(specify)

No abstract available. Title appears to apply to apply to Aim 2 or Aim 3, **OR** can not

Appendix E. Article Inclusion/Exclusion Form

Previewing Only: You cannot submit data from this form
Previewing at Level 3

Refid: 1, Takahashi, T., Saegusa, Y., Takimoto, Y., and Shiragami, M., Pharmacoeconomic analysis of hypertriglyceridemia treatment at the medical institutions, *Yakugaku Zasshi*, 128(12), 2008, p.1783-9
State: Excluded, Level: 1

Submit Data

Does this article APPLY to Aim 2 or Aim 3?

Aim 2. Systems or Guidelines that have been used to evaluate best practices for conducting economic evaluations in health care.

Aim 3. Studies addressing the issues of using economic outcomes in policy and decision making. Include clinical decision making if the decision is not made on the individual clinician level.

DO NOT enter answers for more than one question. This will result in an automatic conflicts. Comments are neutral and can be included with the answer to any of the questions on this form.

1.

Yes, this article **APPLIES** to (*Aim 2 and Aim 3 can be checked if both are applicable*)

Aim 2

Aim 3

2.

No, this article **does not apply** to Aim 2 and Aim 3 for the following reason(s):

Does not apply to economics

Does not apply to health or health care

No original analysis or data (this exclusion includes systematic reviews, commentary, or editorials) Limited case study of a single policy decision: **exclude only** if there is insufficient information about how the economic data was used in decision making process

Study focuses on a single condition: no decision making component, policy making component, or quality of the methods discussion

No cost effectiveness analysis component

Cost-effectiveness analysis only: no decision making component

Study of cost predictors

Costing study

Methods only (descriptive)

Decision making is at the individual clinician level ONLY Other (specify)

Article does not apply (reasons above) BUT is an article of interest and should be pulled for hand searching or background material.

3.

Comments:

Appendix F. Final Article Triage Form

SRS Form

<http://www.srsnexus.com/d2d/ul1/review.asp?mode=previewMode&artic...>

Previewing Only: You cannot submit data from this form



Previewing at Level 5

Refid: 1, Takahashi, T., Saegusa, Y., Takimoto, Y., and Shiragami, M., Pharmacoeconomic analysis of hypertriglyceridemia treatment at the medical institutions, *Yakugaku Zasshi*, 128(12), 2008, p.1783-9
State: Excluded, Level: 1

Submit Data

1. Checklist--level of detail

- Original checklist with generalizable applicability

Aim 2

Final inclusion/exclusion

- Original checklist developed for a specific disease, condition, drug, etc.
 - Partial (original) checklist (containing some elements of economic evaluation)
 - Not an original checklist
- Adapts an original checklist for other

[Clear Selection](#)

2. Exclude for the following reason(s)

- Does not apply to economics
- Does not apply to health or healthcare
- No original data or analysis (systematic reviews, commentary, editorials)
- Limited case study of a single policy decision: exclude only if there is insufficient information about how the economic data was used in decision making process
- Study focuses on a single condition: no decision making component, policy making component, or quality of the methods discussion
- No cost effectiveness analysis component
- Cost-effectiveness analysis only: no decision making component
- Study of cost predictors
- Costing study
- Methods only (descriptive)
- Decision making is at the individual clinician level ONLY
- Article does not apply (reasons above) BUT is an article of interest and should be pulled for hand searching or background

material.

- NOT English--Identify Language
- NOT an AIM 2 article, but does apply to AIM 3

Submit Data

Click a link below to review this article at these other levels.

- [4. Aim 3 Triage](#)
- [6. Aim 3 extraction form](#)
- [7. Aim 2 extraction](#)
- [8. JADAD \(RCT quality\)](#)
- [9. QUALITY--Observational](#)
- [10. QUALITY--Qualitative](#)
- [11. QUALITY--Survey](#)

Form took 0.25 seconds to render

Form Creation Date: Jun 4 2009 9:05AM Form Last Modified: Jun 4 2009 9:27AM

Appendix G. Checklist Details

Table G-1: Checklist described in Adams, 1992¹

<p>Identification of comparison alternatives</p> <ol style="list-style-type: none"> 1. Did the authors provide a detailed description of the alternatives? <ol style="list-style-type: none"> a. Yes, all of the alternatives b. Some, but not all of the alternatives c. No 2. Were any important alternatives omitted? <ol style="list-style-type: none"> a. Yes b. No <p>Identification of costs and monetary savings</p> <ol style="list-style-type: none"> 3. How were direct costs measured? <ol style="list-style-type: none"> a. Opportunity costs b. Charges c. Deflated charges (using cost-to-charge ratios) d. Market prices of resources consumed e. Average cost (per patient day, etc.) f. No details provided g. Other method(s) _____ h. No direct costs measured 4. Were direct incremental costs and monetary savings measured correctly? <ol style="list-style-type: none"> a. Yes, all were measured correctly b. Some, but not all were measured correctly c. No d. Not measured 5. How were overhead costs allocated? <ol style="list-style-type: none"> a. Direct allocation b. Step down allocation c. Step down allocation with iteration d. Simultaneous allocation e. Other method _____ f. Not measured 6. Were the overhead costs allocated appropriately? <ol style="list-style-type: none"> a. Yes, all were allocated appropriately b. Some, but not all were allocated appropriately c. No d. Appropriately not considered e. Not measured 7. Was the opportunity cost of capital measured correctly? <ol style="list-style-type: none"> a. Yes b. No c. Not measured d. Appropriately not considered 8. Was depreciation considered appropriately and measured correctly? <ol style="list-style-type: none"> a. Yes b. No c. Not measured d. Appropriately not considered 9. How were indirect costs measured? <ol style="list-style-type: none"> a. Market valuation 	<p>Identification of benefits</p> <ol style="list-style-type: none"> 13. How were benefits measured? <ol style="list-style-type: none"> a. Life years gained (days, months) b. Quality-adjusted life years (days, months) c. Per positive screening/or correctly diagnosed case d. Change in utilization e. Market valuations f. Opportunity costs g. Compensating variations h. Equivalent variations i. Charges or deflated charges j. Average cost per day (per diem) k. Other method(s) _____ l. Not stated 14. Were direct incremental benefits measured correctly? <ol style="list-style-type: none"> a. Yes, all were measured correctly b. Some, but not all were measured correctly c. No d. Not measured 15. Were indirect incremental benefits measured correctly? <ol style="list-style-type: none"> a. Yes, all were measured correctly b. Some, but not all were measured correctly c. No d. Appropriately not measured e. Not measured 16. Were the incremental benefits adjusted appropriately for differential timing? <ol style="list-style-type: none"> a. Yes, all were adjusted appropriately b. Some, but not all were adjusted appropriately c. No d. Appropriately not adjusted e. Not adjusted 17. Was an appropriate sensitivity analysis conducted correctly? <ol style="list-style-type: none"> a. Complete and correct analysis? b. Partial sensitivity analysis c. No sensitivity analysis <p>Consideration of distributional effect</p> <ol style="list-style-type: none"> 18. Were the distributional effects of the study alternatives considered? <ol style="list-style-type: none"> a. Yes, full discussion b. Yes, partial discussion c. No discussion 19. Were the distributional effects analyzed appropriately? <ol style="list-style-type: none"> a. Yes b. No c. Not analyzed
---	--

<ul style="list-style-type: none"> b. Willingness to pay c. Compensating variations d. Equivalent variations e. Other method(s)_____ f. Not measured <p>10. Were indirect incremental costs and monetary savings measured correctly?</p> <ul style="list-style-type: none"> a. Yes, all were measured correctly b. Some, but not all were measured correctly c. No d. Appropriately not measured e. Not measured <p>11. Were the incremental costs and monetary savings (direct and indirect) adjusted appropriately for differential timing (discounting, inflation and time-value of money)?</p> <ul style="list-style-type: none"> a. Yes, all were adjusted appropriately b. Some, but not all were adjusted appropriately c. No d. Appropriately not measured e. Not adjusted <p>12. Was an appropriate sensitivity analysis conducted correctly?</p> <ul style="list-style-type: none"> a. Complete and correct analysis b. Partial sensitivity analysis c. No sensitivity analysis 	<p>Aggregation of valued costs and effects</p> <p>20. Were the incremental costs and benefits aggregated correctly?</p> <ul style="list-style-type: none"> a. Yes b. Only partial c. No <p>21. Were the uncertainties regarding the cost and benefit estimates, the incidence of side effects, probabilities of being in different states of nature, and the clinical decision making process discussed?</p> <ul style="list-style-type: none"> a. Full discussion b. Partial discussion c. No discussion
---	--

Table G-2: Checklist described in Gerard, 1992²

<p>Background characteristics</p> <ol style="list-style-type: none"> 1. Type of health care strategy 2. Type of disease 3. Type of health care intervention 4. client group 5. Country of study 6. Academic Departments . Research Centres 7. Person(s) most responsible for evaluation/studies with/without a health economics input 8. Funding agency <p>Technical criteria</p> <ol style="list-style-type: none"> 9. Objectives of study* 10. Type and adequacy of comparisons* <p>Costs</p> <ol style="list-style-type: none"> 11. Cost perspective 12. Cost coverage 13. Cost comprehensiveness* 14. Clarity of cost measurement 15. Type of cost measurement 16. Source and description of cost valuation 17. Clarity of the price base used 	<p>Outcome</p> <ol style="list-style-type: none"> 18. Outcome coverage* 19. Type of outcome coverage 20. Clarity and type of added life measurement 21. Type of quality of life measurement 22. Quality of life measurement techniques 23. Source of QALY valuations 24. Clarity of assignment of valuation in health states 25. Discounting 26. Sensitivity analysis* 27. Source/type of effectiveness evidence 28. timing of CUA with effectiveness evidence 29. Stage of technical development of intervention <p>Policy relevance criteria</p> <ol style="list-style-type: none"> 30. Additional value of CUA <i>vis-à-vis</i> CEA 31. Results brought to policy-makers' attention 32. Generalise results to other settings 33. Use and construction of QALY league tables <p>General quality of studies</p> <ol style="list-style-type: none"> 34. Sensible objectives 35. Main potential value 36. Execution 37. General (or overall) assessment <p>* denotes key technical criteria used in the overall assessment of the quality of the study. This assessment of overall technical quality is applied only to studies judged worthwhile because objectives were sensible and the CUA approach was appropriate.</p>
--	--

Table G-3: Checklist described by Sacristan, 1993³

<p>1. Definition of study aim Does a well defined question exist? Are the perspectives and alternatives compared clearly specified?</p> <p>2. Sample selection Are the types of patients chosen suitable and are they specified? Are the diagnostic criteria adequately specified?</p> <p>3. Analysis of alternatives Are all the relevant alternatives analyzed? Are the comparison alternatives suitable? Is this the most commonly used treatment? Is the indication the most relevant one? Are adequate dosages used? Are the treatments reproducible? (doses, interval, duration, etc.) Is the 'do nothing' option suitable or should it be analyzed? Is a decision analysis applied?</p> <p>4. Analysis of perspective Is it clearly specified (society, patient, hospital, etc.)? Is it justified for the question asked?</p> <p>5. Measurement of benefits Is it adequate for the question asked and the perspective? Are the data on the effectiveness of alternatives adequately established? Is the main assessment variable, objective and relevant? Is the time fixed for the evaluation sufficient and is it specified? Are the results quantified by time?</p> <p>6. Measurement of costs Is the measurement of costs suitable for the perspective Are the costs up to date and the prices those of the market? Is an adjustment of future costs and benefits performed?</p>	<p>7. Is the type of analysis used suitable? Financial terms: cost-benefit Physical "units: cost-effectiveness Quality of life/utility: cost-utility Equal benefits: cost-minimization</p> <p>8. Analysis of the results If intermediate variables are used, are they representative of the end benefit? Is a marginal analysis performed? Are the costs and consequences of adverse affects analyzed?</p> <p>9. Is the evaluation suitable if carried out within a clinical trial? Is the suitable methodology employed? Are the statistical methods used adequate? Is an analysis according to 'intention to treat" made? Are costs resulting from the trial, which differ from those in normal practice, taken into account?</p> <p>10. Are the assumptions and the limitations of the study discussed? Is a sensitivity analysis performed? Do the assumptions have a basis? Is the exclusion of any important variable analyzed or justified? If intermediate endpoints are assumed, are limitations discussed?</p> <p>11. Are the possible ethical problems discussed and identified?</p> <p>12. Conclusions Are they justified? Can they be generalized? Can they be extrapolated to daily clinical practice?</p>
---	--

Table G-4: Checklist described in Clemens, 1995⁴

<p>Research design</p> <ol style="list-style-type: none"> 1. Individual researchers should be free to choose from multiple research design and data source options that have been described in the literature. <p>Protocol/report</p> <ol style="list-style-type: none"> 2. Study objectives and rationale for the outcome measures, data sources and analysis methods chosen should be stated in a research proposal or protocol. 3. A final report should be available that describes all assumptions, methods and data sources in sufficient detail that readers can reproduce the key results. <p>Contents of protocol/report</p> <ol style="list-style-type: none"> 4. The study should clearly specify the target population or patient subgroups. 5. Alternative treatments such as other drugs, surgical intervention or 'do nothing' should be stated in the study design including a discussion of their costs, frequency of use and relevance for the population of interest. 6. Time horizons for the analysis should be stated and should be based on the likely use and effect of the drug, together with the weight and strength of data availability. 7. The study should clearly define the type of analyses and rationale for their use. 8. Uncertainty should be demonstrated by: (i) statistical analysis to address random events; and (ii) sensitivity analysis to cover a range of assumptions. <p>Costs/resources</p> <ol style="list-style-type: none"> 9. The perspective of the analysis (societal, third-party payer, etc) and clear reasons for its selection should be stated. 10. Resources should be relevant to the perspective and described in natural units before a cost analysis is performed. 	<ol style="list-style-type: none"> 11. Categories of costs (direct medical, direct nonmedical and indirect) should be identified and presented separately. 12. Resources/costs should be presented as increments (difference between alternatives) as well as totals. 13. Costs should be discounted for analyses with time horizons greater than 1 year. 14. The source(s) and methods of deriving the costs/charges should be clearly stated and validated. <p>Effectiveness/benefits</p> <ol style="list-style-type: none"> 15. Consequences being evaluated (monetary benefit, effectiveness, quality of life, utility, efficacy, safety, morbidity, mortality) and their sources should be clearly stated. 16. Effectiveness/benefit measures should be discounted, when appropriate, for analyses with time horizons of greater than 1 year. 17. Effectiveness/benefit measures should be presented as increments (difference between alternatives) as well as totals. <p>Data sources</p> <ol style="list-style-type: none"> 18. Methods for the estimation of quantities and unit costs are described. 19. The quality of the database should be documented (e.g., methods of data assembly, limitations such as missing data, validation methods). 20. Expert clinical opinion may be used for certain limited purposes when other data sources are not available (e.g., dealing with missing data, adjusting for protocol-driven resource usage, estimating probabilities in decision analysis, and when collection of primary data is unethical). <p>Extrapolation of results to other settings</p> <p>Effectiveness and resource data from one setting may be extrapolated to other settings if differences in medical intervention, treatment patterns, outcomes of existing treatment, and populations are addressed.</p>
--	--

Table G-5: Checklist described by the US Panel, 1996⁵⁻⁷

<p>Framework</p> <ol style="list-style-type: none"> 1. Background of the problem 2. General framing and design of the analysis 3. Target population for intervention 4. Other program descriptors (eg, care setting, model of delivery, timing of intervention) 5. Description of comparator programs 6. Boundaries of the analysis 7. Time horizon 8. Statement of the perspective of the analysis <p>Data and Methods</p> <ol style="list-style-type: none"> 9. Description of event pathway 10. Identification of outcomes of interest in analysis 11. Description of model used 12. Modeling assumptions 13. Diagram of event pathway 14. Software used 15. Complete description of estimates of effectiveness, resource use, unit costs, health states, and quality of life weights and their sources 16. Methods for obtaining estimates of effectiveness, costs and preferences 17. Critique of data quality 18. Statement of year of costs 19. Statement of method used to adjust costs for inflation 20. Statement of type of currency 21. Source and methods for obtaining expert judgment 22. Statement of discount rates 	<p>Results</p> <ol style="list-style-type: none"> 23. Results of model validation 24. Reference case results (discounted at 3% and undiscounted): total costs and effectiveness, incremental costs and effectiveness, and incremental cost effectiveness ratios 25. Results of sensitivity analyses 26. Other estimates of uncertainty, if available 27. Aggregate cost and effectiveness information 28. Disaggregated results, as relevant 29. Secondary analyses using 5% discount rate 30. Other secondary analyses, as relevant <p>Discussion</p> <ol style="list-style-type: none"> 31. Summary of reference case results 32. Summary of sensitivity of results to assumptions and uncertainties in the analysis 33. Discussion of analysis assumptions having important ethical implications 34. Limitations of the study 35. Relevance of study results for specific policy questions or decisions 36. Results of related cost-effectiveness analyses 37. Distributive implications of an intervention
--	---

Table G-6: The Paediatric Quality Appraisal Questionnaire Described by Ungar, 2003⁸

<p>Economic Evaluation</p> <ol style="list-style-type: none"> 1. Is the research question posed in terms of costs and consequences? 2. Is a specific type of economic analysis technique performed? 3. What type of analytic technique is performed, according to the authors? <p>Comparators</p> <ol style="list-style-type: none"> 4. Is there a rationale for choosing the intervention(s) being investigated? 5. Is there a rationale for choosing the alternative program(s) or intervention(s) used for comparison? 6. Does the report describe the alternatives in adequate detail? 7. Is a description of the event pathway provided? 8. Is a formal decision analysis performed? <p>Target Population</p> <ol style="list-style-type: none"> 9. Is the target population for the intervention identified? 10. Are the subjects representative of the population to which the intervention is targeted? <p>Time Horizon</p> <ol style="list-style-type: none"> 11. Is there a time horizon for both costs and outcomes? 12. Do the authors justify the time horizon selected? <p>Perspective</p> <ol style="list-style-type: none"> 13. Is a perspective for the analysis given? 14. Is a societal perspective taken, either alone or in addition to other perspectives? 15. When there is more than one perspective, are the results of each perspective presented separately? <p>Costs and Resource Use</p> <ol style="list-style-type: none"> 16. Are all relevant costs for each alternative included? 17. Are opportunity costs of lost time (productivity costs) for parents and informal caregivers measured when required? 18. Do cost item identification and valuation extend beyond the health-care system to include school and community resources when necessary? 19. Are future salary and productivity changes of the child taken into consideration when appropriate? 20. Are all of the sources for estimating the volume of resource use described? 21. Are all the sources for estimating all of the unit costs described? <p>Outcomes</p> <ol style="list-style-type: none"> 22. Is a primary health outcome given? 23. Do the authors justify the health outcome(s) selected? 24. Is effectiveness, rather than efficacy, assessed? 25. What approach is used to assess the effectiveness/efficacy? 26. Are the details of the design of the effectiveness/efficacy study(s) provided? 27. Are the results of the efficacy/effectiveness of alternatives reported? 28. Are school/day-care absences taken into consideration? 29. If intermediate outcome variables are used, are they linked by evidence or reference to the end benefit? 	<p>Quality of Life</p> <ol style="list-style-type: none"> 30. If quality of life is measured, what type of instrument is used? 31. Whose quality of life is assessed? 32. Who performed the quality-of-life assessment? <p>Analysis</p> <ol style="list-style-type: none"> 33. Are costs AND outcomes measured in units appropriate for the indicated analytic technique? 34. For prospective studies that use interviews, questionnaires, or surveys, how are data obtained in studies involving young children? 35. How are direct costs valued? 36. How are productivity costs valued? 37. Are costs valued appropriately? 38. Is the valuation of outcomes appropriate for the type of analysis? 39. What is the unit of analysis used for expressing the <i>final</i> results? 40. Are quantities of resources used reported separately from their unit costs? 41. Are the costs aggregated correctly? 42. Are details of statistical tests and confidence intervals given for stochastic data? <p>Discounting</p> <ol style="list-style-type: none"> 43. When required, are costs and consequences that occur over more than 1 year discounted to their present values? 44. If costs or benefits are not discounted when the time horizon exceeds 1 year, is an explanation provided? <p>Incremental Analysis</p> <ol style="list-style-type: none"> 45. Are incremental estimates of costs and outcomes presented? 46. Are the incremental estimates summarized as incremental ratios? 47. Are confidence intervals/limits calculated for incremental ratios or incremental estimates of costs and outcomes? <p>Sensitivity Analysis</p> <ol style="list-style-type: none"> 48. Are all important assumptions given? 49. Is a sensitivity analysis performed? 50. Do the authors justify the alternative values or ranges for sensitivity analysis? 51. What methods are used to assess uncertainty? <p>Conflict of Interest</p> <ol style="list-style-type: none"> 52. Does the article present the relationship with the sponsor of the study? 53. Does the article indicate that the authors had independent control over the methods and right to publish? <p>Conclusions</p> <ol style="list-style-type: none"> 54. Is the answer to the study question provided? 55. Are the most important limitations of the study discussed? 56. Do the authors generalize the conclusions to other settings or patient/client groups? 57. Global impression of the quality of the article.
--	--

Table G-7: The British Medical Journal Checklist described by Drummond, 1996⁹

<p>Study design</p> <ol style="list-style-type: none"> 1. The research question is stated 2. The economic importance of the research question is stated 3. The viewpoints of the analysis are clearly stated and justified 4. The rationale for choosing the alternative programmes or interventions compared is stated 5. The alternatives being compared are clearly described 6. The form of economic evaluation used is stated 7. The choice of form of economic evaluation is justified in relation to the questions addressed <p>Data collection</p> <ol style="list-style-type: none"> 8. The sources of effectiveness estimates used are stated 9. Details of the design and results of effectiveness study are given (if based on a single study) 10. Details of the method of synthesis or meta-analysis of estimates are given (if based on an overview of a number of effectiveness studies) 11. The primary outcome measure(s) for the economic evaluation are clearly stated 12. Methods to value health states and other benefits are stated 13. Details of the subjects from whom evaluations were obtained are given 14. Productivity changes (if included) are reported separately 15. The relevance of productivity changes to the study question is discussed 16. Quantities of resources are reported separately from their unit costs 17. Methods for the estimation of quantities and unit costs are described 18. Currency and price data are recorded 19. Details of currency or price adjustments for inflation or currency conversion are given 20. Details of any model used are given 21. The choice of model used and the key parameters on which it is based are justified 	<p>Analysis and interpretation of results</p> <ol style="list-style-type: none"> 22. Time horizon of costs and benefits is stated 23. The discount rate(s) is stated 24. The choice of rate(s) is justified 25. An explanation is given if costs or benefits are not discounted 26. Details of statistical tests and confidence intervals are given for stochastic data 27. The approach to sensitivity analysis is given 28. The choice of variables for sensitivity analysis is justified 29. The ranges over which the variables are varied are stated 30. Relevant alternatives are compared 31. Incremental analysis is reported 32. Major outcomes are presented in a disaggregated as well as aggregated form 33. The answer to the study question is given 34. Conclusion follow from the data reported Conclusions are accompanied by the appropriate caveats
--	--

Table G-8: The Quality of Health Economic Studies Instrument described by Chiou, 2003¹⁰

Item	Weightings
1. Was the study objective presented in a clear, specific, and measurable manner?	7
2. Were the perspective of the analysis (societal, third-party payer, and so on) and reasons for its selection stated?	4
3. Were variable estimates used in the analysis from the best available source (i.e., Randomized Control Trial-Best, Expert Opinion-Worst)?	8
4. If estimates came from a subgroup analysis, were the groups pre-specified at the beginning of the study?	1
5. Was uncertainty handled by: (i) statistical analysis to address random events; (ii) sensitivity analysis to cover a range of assumptions?	9
6. Was incremental analysis performed between alternatives for resources and costs?	6
7. Was the methodology for data abstraction (including value health states and other benefits) stated?	5
8. Did the analytic horizon allow time for all relevant and important outcomes? Were benefits and costs that went beyond 1 year discounted (3–5%) and justification given for the discount rate?	7
9. Was the measurement of costs appropriate and the methodology for the estimation of quantities and unit costs clearly described?	8
10. Was the primary outcome measure(s) for the economic evaluation clearly stated and were the major short term, long term and negative outcomes included?	6
11. Were the health outcomes measures/scales valid and reliable? If previously tested valid and reliable measures were not available, was justification given for the measures/scales used?	7
12. Were the economic model (including structure), study methods and analysis, and the components of the numerator and denominator displayed in a clear transparent manner?	8
13. Were the choice of economic model, main assumptions and limitations of the study stated and justified?	7
14. Did the author(s) explicitly discuss direction and magnitude of potential biases?	6
15. Were the conclusions/recommendations of the study justified and based on the study results?	8
16. Was there a statement disclosing the source of funding for the study?	3

Table G-9: The Consensus Health Economic Criteria List described by Evers, 2005¹¹

<ol style="list-style-type: none"> 1. Is the study population clearly described? 2. Are competing alternatives clearly described? 3. Is a well-defined research question posed in answerable form? 4. Is the economic study design appropriate to the stated objective? 5. Is the chosen time horizon appropriate in order to include relevant costs and consequences? 6. Is the actual perspective chosen appropriate? 7. Are all important and relevant costs for each alternative identified? 8. Are all costs measured appropriately in physical units? 9. Are costs valued appropriately? 10. Are all important and relevant outcomes for each alternative identified? 	<ol style="list-style-type: none"> 11. Are all outcomes measured appropriately? 12. Are outcomes valued appropriately? 13. Is an incremental analysis of costs and outcomes of alternatives performed? 14. Are all future costs and outcomes discounted appropriately? 15. Are all important variables, whose values are uncertain, appropriately subjected to sensitivity analysis? 16. Do the conclusions follow from the data reported? 17. Does the study discuss the generalizability of the results to other settings and patient/client groups? 18. Does the article indicate that there is no potential conflict of interest of study researcher(s) and funder(s)? 19. Are ethical and distributional issues discussed appropriately?
---	--

Table G-10. Checklist to Frame Health Technology Assessments, Grutters, 2011¹²

1. Objective How will the Health Technology Assessment be used?	7. Time horizon Which time horizon is relevant for the decision problem?
2. Audience What is the audience (principal users) for the HTA?	8. Consequences Which consequences are relevant for the decision problem?
3. Perspective Which viewpoint or perspective is relevant for the HTA?	9. Patient Use What is the patient use that is relevant for the decision problem?
4. Population What is the patient population relevant for the decision problem?	10. Professional Use What is the use of the technology by health care professionals that is relevant for the problem?
5. Comparators What are relevant comparators for the decision problem?	11. Price and Resource Use What price level and resource use are relevant for the decision problem?
6. Clinical Practice How are the technologies embedded in clinical practice?	

References

- Adams ME, McCall NT, Gray DT, Orza MJ, Chalmers TC. Economic analysis in randomized control trials. *Med Care* 1992; 30(3):231-43.
- Gerard K. Cost-utility in practice: a policy maker's guide to the state of the art. *Health Policy* 1992; 21(3):249-79.
- Sacristan JA, Soto J, Galende I. Evaluation of pharmacoeconomic studies: utilization of a checklist. *Ann Pharmacother* 1993; 27(9):1126-33.
- Clemens K, Townsend R, Luscombe F, Mauskopf J, Osterhaus J, Bobula J. Methodological and conduct principles for pharmacoeconomic research. *Pharmaceutical Research and Manufacturers of America. Pharmacoeconomics* 1995; 8(2):169-74.
- Siegel JE, Weinstein MC, Russell LB, Gold MR. Recommendations for reporting cost-effectiveness analyses. Panel on Cost-Effectiveness in Health and Medicine. *JAMA* 1996; 276(16):1339-41.
- Weinstein MC, Siegel JE, Gold MR, Kamlet MS, Russell LB. Recommendations of the Panel on Cost-effectiveness in Health and Medicine. *JAMA* 1996; 276(15):1253-8.
- Russell LB, Gold MR, Siegel JE, Daniels N, Weinstein MC. The role of cost-effectiveness analysis in health and medicine. Panel on Cost-Effectiveness in Health and Medicine. *JAMA* 1996; 276(14):1172-7.
- Ungar WJ, Santos MT. The Pediatric Quality Appraisal Questionnaire: an instrument for evaluation of the pediatric health economics literature. *Value Health* 2003; 6(5):584-94.
- Drummond MF, Jefferson TO. Guidelines for authors and peer reviewers of economic submissions to the BMJ. The BMJ Economic Evaluation Working Party. *BMJ* 1996; 313(7052):275-83.
- Chiou CF, Hay JW, Wallace JF *et al.* Development and validation of a grading system for the quality of cost-effectiveness studies. *Med Care* 2003; 41(1):32-44.
- Evers S, Goossens M, de Vet H, van Tulder M, Ament A. Criteria list for assessment of methodological quality of economic evaluations: Consensus on Health Economic Criteria. *Int J Technol Assess Health Care* 2005; 21(2):240-5.
- Grutters JP, Seferina SC, Tjan-Heijnen VC, van Kampen RJ, Goettsch WG, Joore MA. Bridging trial and decision: a checklist to frame health technology assessments for resource allocation decisions. *Value Health* 2011; 14(5):777-84.