Methods Research Report

Methods for Benefit and Harm Assessment in Systematic Reviews

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

Methods Research Report

Methods for Benefit and Harm Assessment in Systematic Reviews

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

Contract No. 290-2007-10061-I

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

Investigators
Cynthia M. Boyd, M.D., M.P.H.
Sonal Singh, M.D., M.P.H.
Ravi Varadhan, Ph.D.
Carlos O. Weiss, M.D., M.H.S.
Ritu Sharma, B.Sc.
Eric B. Bass, M.D., M.P.H.
Milo A. Puhan, M.D., Ph.D.

AHRQ Publication No. 12(13)-EHC150-EF
November 2012

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 author(s), 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.

The investigators have no relevant financial interests in the report. The investigators have no employment, consultancies, honoraria, or stock ownership or options, or royalties from any organization or entity with a financial interest or financial conflict with the subject matter discussed in the report.

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

Suggested citation: Boyd CM, Singh S, Varadhan R, Weiss CO, Sharma R, Bass EB, Puhan MA. Methods for Benefit and Harm Assessment in Systematic Reviews. Methods Research Report. (Prepared by the Johns Hopkins University Evidence-based Practice Center under contract No. 290-2007-10061-I). AHRQ Publication No. 12(13)-EHC150-EF. Rockville, MD: Agency for Healthcare Research and Quality; November 2012.

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 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.gov.

Carolyn M. Clancy, M.D.
Director
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

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

Parivash Nourjah, Ph.D.
Task Order Officer
Agency for Healthcare Research and Quality

Acknowledgments

The team would like to thank Daniela Puhan-Vollenweider, M.D., Manisha Reuben, and Renee Wilson for their assistance with the formatting of this report and Eric Vohr and Faye Rivkin for their editing and Dr. Parivash Nourjah for her valuable insight throughout the project. We also thank Dr. Kevin Frick for his valuable comments on a draft of the report.

Thank you to the following Technical Expert Panel members for their time and expert guidance:

Elie Akl, M.D.
University of New York
Buffalo, NY

Marguerite A. Koster, M.A., M.F.T.,
Kaiser Permanente, Southern California

Robert L. Kane, M.D.
University of Minnesota
Minneapolis, MN

Thomas A. Trikalinos, M.D., Ph.D.
Tufts Medical Center
Boston, MA

Wiley Chan, M.D.
Northwest Permanente

Thank you to the following Peer Reviewers for their time and expert guidance:

Jonathan Treadwell, Ph.D.
Evidence-based Practice Center, ECRI Institute,
Plymouth Meeting, PA

Robert L. Kane, M.D.
University of Minnesota
Minneapolis, MN

Joshua Cohen, Ph.D.
Tufts University
Boston, MA

Melissa McPheeters, Ph.D., M.P.H.
Vanderbilt University
Nashville, TN

Francois Sainfort, Ph.D.
University of Minnesota
Minneapolis, MN

Methods for Benefit and Harm Assessment in Systematic Reviews

Structured Abstract

Introduction. Systematic reviewers are challenged by how to report and synthesize information about benefits and harms of medical interventions so that decisionmakers with varying preferences can better assess the balance of benefit and harm. Quantitative approaches exist for assessing benefits and harms, but it is unclear whether they are applicable to systematic reviews.

Objectives. The objectives of this report are: (1) to describe the challenges of quantitative approaches for assessing benefits and harms, (2) to describe methodological characteristics of existing quantitative approaches for assessing benefits and harms, (3) to determine the role of values and preferences in assessing benefits and harms across each step of a systematic review and (4) to formulate principles for assessing benefits and harms in systematic reviews.

Process. We formed a multidisciplinary team with expertise in clinical medicine, systematic reviews, statistics, and epidemiology**.** The team reviewed the literature on quantitative approaches for assessing benefits and harms of medical interventions, and held 12 weekly meetings to establish consensus about: 1) the challenges in assessing benefits and harms; 2) the methodological characteristics of approaches that have been used; and 3) the role of values and preferences when assessing benefits and harms in systematic reviews.

The team used that information to formulate principles for analyzing benefits and harms in systematic reviews so that decisionmakers are able to weigh the benefits and harms for a given population. An external panel of experts provided input in this process.

Results. Our team identified numerous challenges for the assessment of benefits and harms. The main challenges relate to selection of health outcomes important to patients, information asymmetry (e.g., reliable and robust data on benefits with sparse data on harms), and calculation of statistical uncertainty if benefit and harm are put on the same scale using a benefit harm comparison metric, and consideration of patient preferences.

We identified 16 quantitative approaches for the assessment of benefits and harms. Twelve of the methods can be used in a systematic review because the methods can be applied with the types of summary data that are typically reported and do not require individual patient data. Simpler approaches, such as the ratio of the number needed to treat to the number needed to harm, may be suitable for relatively simple decisionmaking contexts where relevant benefit and harm outcomes are few in number and similar in importance. More complex approaches are needed for decisionmaking contexts having a large number of relevant benefits and harms.

For individual-level decisions, values and preferences are key for determining the balance of benefit and harm. Choices are made by decisionmakers that are informed by the preferences of patients and other considerations. These choices, and therefore preferences, have an important

role in determining how benefits and harms are assessed in systematic reviews. These choices and preferences also affect how guideline developers frame recommendations, how regulatory bodies make decisions at the population level, and how clinicians, patients, and other end users make decisions at the individual level.

The team formulated principles to conduct comparative assessments of benefits and harms in the context of a systematic review. For example, we recommend that systematic reviews define the decisionmaking context, report the sources of evidence used (e.g., estimates of baseline risks or treatment effects), be explicit about if and how patient preferences are considered, and provide a rationale for choosing a particular quantitative approach for comparative assessment of benefits and harms.

Conclusion. Quantitative approaches for comparative assessment of benefits and harms have strengths and limitations. The choice of a particular approach depends on the decisionmaking context, the quality and quantity of available data, and the epidemiological-statistical expertise of the systematic review team. A quantitative approach may help to improve the transparency of a review, relative to a qualitative approach, by being explicit about how benefits and harms are estimated and compared. Such transparency may help decisionmakers give proper consideration to complex information about benefits and harms.

Contents

Tables

Figures

Introduction

Systematic reviews assess the comparative effectiveness and safety of health care interventions and are useful to a variety of decisionmakers. The reviews usually consider a range of outcomes that are relevant to patients and other stakeholders, and often include a variety of study designs such as trials and observational studies. Systematic reviews also attempt to identify subgroups (e.g., elderly and ethnic minorities) for which the effectiveness and harms of interventions may vary.

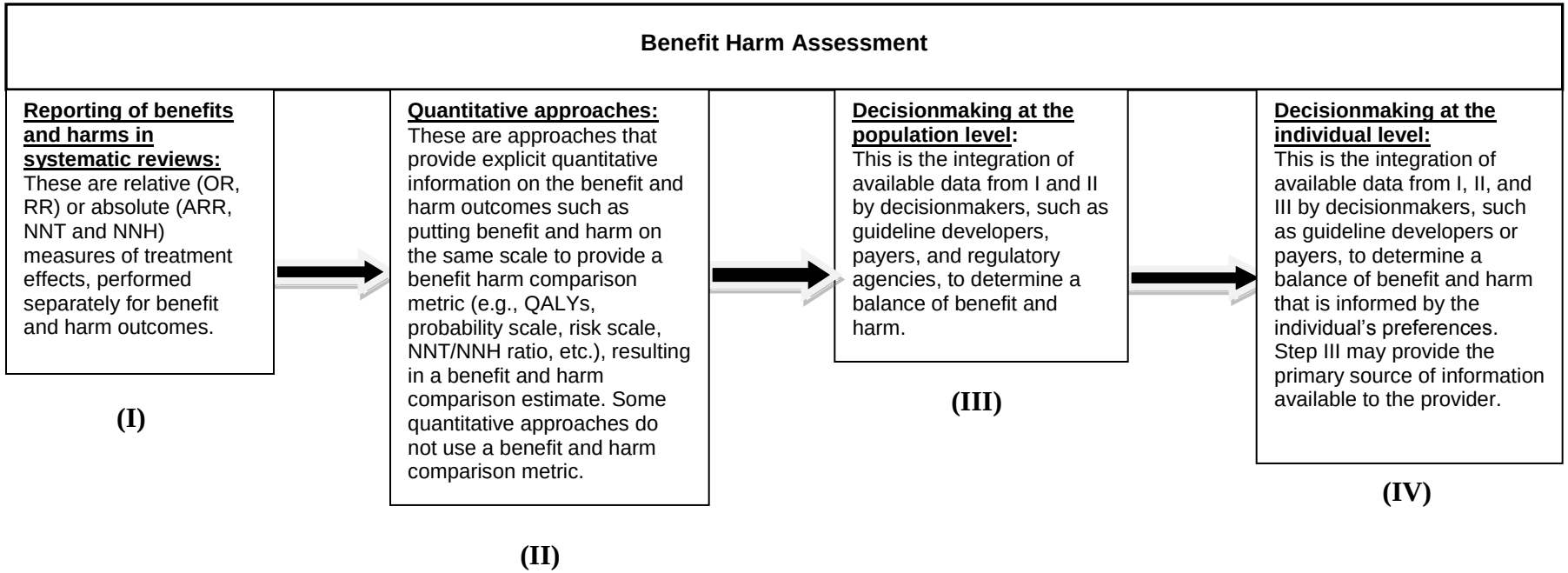
In certain systematic reviews, reviewers may assess benefits and harms separately, with separate key questions.[1,2] Reviewers may report these results in separate tables and in separate sections of the report.[3-5] The metrics used to report benefit or harm outcomes across studies are usually mean differences between study groups, odds ratios, or relative risks. Systematic reviews also tend to describe the strength (i.e., quality) of the evidence for each benefit and harm outcome separately.[1] Often, the information is asymmetric, with more reliable and robust data on benefits as compared with harms. A recent review reported that only 19 percent of 104 comparative effectiveness studies of medications primarily focused on harm outcomes.[6]

Quantitative approaches exist to estimate the balance of benefits and harms to better inform medical and public health decisions made by stakeholders (e.g., patients or healthy subjects, health care and public health providers, payers, policymakers, or regulatory agencies such as the Food and Drug Administration or European Medicines Agency). Quantitative approaches use formulas or statistical graphs to compare benefits and harms, expressed by metrics such as an odds ratio, relative risk, or absolute difference. These quantitative approaches can combine benefit and harm outcomes on a single scale or report them separately. Some quantitative approaches also incorporate quantitative information about patient preferences. However, these approaches have primarily focused on individual studies, and some have relied on information collected from individual study participants that is unavailable to most systematic reviewers. As such, researchers have not rigorously applied these approaches to systematic reviews that synthesize information across multiple studies.

Several methodological issues are inherent to using these quantitative approaches to assess benefits and harms in systematic reviews. Systematic reviewers have rarely applied these approaches and lack guidance on how to use the approaches in systematic reviews.

Figure 1 illustrates how quantitative approaches may fit into the process of assessing benefits and harms. The process of assessing benefits and harms relies on several steps:

Figure 1. Stages of benefit and harm assessment



ARR = absolute risk reduction; NNT = number needed to treat; NNH = number needed to harm OR = odds ratio; QALY = quality-adjusted life years; RR = relative risk

Step I: Reporting of benefits and harms in absolute or relative terms in systematic reviews or meta-analysis;

Step II: Use of quantitative approaches that provide explicit quantitative information on the benefit and harm outcomes and that may compare benefits and harms (or justification for not using a quantitative approach);

Step III: Judgments made by decisionmakers* at the population level after an assessment of the balance of benefits and harms; and

Step IV: Shared decisions made by providers and patients at the individual level incorporating their values and preferences, often using information from decisions at the population level.

While Evidence-based Practice Centers generally do not weigh benefits against harms when conducting systematic reviews for the Agency for Healthcare Research and Quality (AHRQ), they need to report them in a manner that makes it possible to do so. Decisionmakers need to apply valid methods to use the information in the systematic reviews for health care decisionmaking.

The specific objectives of this project were to:

1. Describe the challenges of quantitative approaches for assessing benefits and harms in systematic reviews;
2. Describe methodological characteristics of existing quantitative approaches for assessing benefits and harms in systematic reviews;
3. Determine the role of values and preferences in assessing benefits and harms across each step of a systematic review (very simply, values refer to general dispositions, and preferences refer to degrees of desirability that people associated with a health state.[7] We provide more detailed definitions in the section: Influence of Values and Preferences in Assessing Benefits and Harms); and
4. Formulate principles for assessing benefits and harms in systematic reviews so that decisionmakers are better able to weigh the benefits and harms (including adverse effects and burdens) for a population and for subgroups for which this balance may vary, after accounting for values and preferences.

Our scope was limited to the assessment of quantitative approaches for assessing benefits and harms in the context of evidence synthesis. We reviewed quantitative approaches that provide an assessment of the benefits and harms of health care interventions, not diagnostic tests. Theoretical and qualitative approaches are beyond the scope of this report, as are detailed methods and guidance on the conduct of systematic reviews.[8,9]

* Step III involves different types of decisionmakers that may use systematic reviews to inform their decisions (e.g., guideline panels and health care payers). Depending on the decisionmaker, patient-level characteristics may require attention during population-level decisions. The choices made by these decisionmakers are likely to vary.

Methods

Our team, consisting of clinicians, epidemiologists, and statisticians, first addressed the first two objectives of this project: (1) to identify challenges for assessing benefits and harms in systematic reviews, and (2) to describe methodological characteristics of existing quantitative approaches for assessing benefits and harms in systematic reviews.

We selected a sample of evidence reports published between 2007 and 2011 from the Effective Health Care Web site of the Agency for Healthcare Research and Quality (AHRQ),[3 4] and two recent systematic reviews that reported on benefits and harms.[5,10-12] The team then searched for quantitative approaches for assessing benefits and harms, examining key articles garnered from the team's reference libraries, including prior work on methods for describing benefits and harms. We did not perform a formal systematic review of the literature because a review on the topic of benefit harm assessment already existed, and because our focus was on organizing available approaches in the context of systematic reviews.[13-17]

We capitalized on the work done previously and created a list of relevant approaches, which allowed us to concentrate on the main focus of synthesizing information about available methods for use in systematic reviews.[13] The team looked for articles that quantitatively assessed at least one outcome each for both benefit and harm of a medical or public health intervention and included approaches that analyzed benefit and harm outcomes separately, as well as approaches that provided a benefit and harm comparison metric for the balance of benefit and harm (e.g., ratio of number needed to treat [NNT] to number needed to harm [NNH]). The team also evaluated articles citing any of the above articles to determine whether they described additional relevant concepts or quantitative approaches. Finally, we screened the reference lists of all included articles for more relevant articles.

One team member led the discussion of challenges, the methodological characteristics of quantitative approaches, values and preferences, and principles. All team members were involved in the weekly discussion. Subsequently, we circulated notes from the meeting until saturation of themes and consensus was achieved on operational terms. Our team discussed the existing quantitative approaches for benefit harm assessment in 12 1-hour sessions.

The discussion helped us to define the properties that characterize quantitative approaches for assessing benefits and harms. To develop an organizing framework for quantitative approaches, the team iteratively defined a list of key characteristics to describe existing quantitative approaches for assessing benefits and harms and that allowed comparisons across the quantitative approaches. For example, the team identified the types of data needed for each quantitative approach, the assumptions underlying the quantitative approaches, the benefit and harm comparison metrics used, and the way researchers communicate the results of the benefit harm assessment to decisionmakers. In addition, the team defined desired properties of quantitative approaches for benefit and harm assessment that were beyond statistical considerations, such as the populations considered, the comprehensiveness and quality of data collected, and the sources of evidence considered. The team also recognized and recorded limitations inherent in each of the quantitative approaches that could threaten their usefulness. We described all quantitative approaches using the final list of key characteristics and unified the approaches for comparison in a table.

The team shared these key characteristics with a diverse panel of external experts (the Technical Expert Panel). The expert panel was composed of systematic reviewers, experts in

patient preferences and members of other Evidence Based Practice Centers, including end-users of systematic reviews.

Starting with the list of key characteristics, the team identified the properties a quantitative approach for benefit harm assessment should possess to be valid and relevant to decisionmakers. In doing so, the team considered the advantages and disadvantages of existing quantitative approaches and how the approaches could be refined to have as many desirable properties as possible.

For Objective 3, our team considered the role of values and patient preferences in systematic reviews and quantitative approaches to assessing benefits and harms. We devoted three additional weekly sessions to evaluating the implicit and explicit role of values and preferences during the process of evidence generation and evidence synthesis. These included evaluating the role of values and preferences in moving from evidence generation to development of evidence-based medicine tools. We evaluated how choices of investigators, systematic reviewers, and policymakers may impact the assessment of benefits and harms in systematic reviews.

The team reviewed the results for Objectives 1 through 3 and developed principles for assessing the balance of benefits and harms in systematic reviews (Objective 4). To guide systematic reviewers who plan to conduct or inform a quantitative benefit harm assessment, these principles addressed various stages of the systematic review process. The team revised these preliminary principles after electronic review and input from the Technical Expert Panel.

Challenges of Quantitative Approaches for Assessing Benefit and Harm

The challenges in assessing benefits or harms in systematic reviews are listed below and shown in Table 1. The team organized its listing of challenges around the Population, Intervention, Comparator, Outcome, Timeframe, and Setting (PICOTS) criteria,[9] with some minor adaptations to incorporate the challenges of preferences.

Table 1. Summary listing of challenges in assessing benefits and harms in systematic reviews

Criteria	Challenges
Populations	Benefits and harms, or uncertainty about the balance between them, may vary for subgroups of the population.
Interventions	Benefits and harms may vary due to differences in the fidelity of interventions across studies.
Comparisons	The estimates of effect from a study may reflect the beneficial effect of one intervention or the harmful effect of a comparator.
Outcomes	Assessment may depend on the linkages between surrogates and health outcomes for specific interventions and whether there is variation in these linkages across subgroups. Assessment may also depend on composite outcome measures, where individual elements of the composite measure may have different effect sizes and a different gradient of preferences.
Time Horizon	The time horizon in studies may be inadequate for assessing all benefits and harms (e.g., early benefit, late risks)
Study Designs	Studies may be designed to provide more robust data on benefits than on harms. Assessment of harms may require study designs other than RCTs, such as observational studies or case reports.
Strength of Evidence	The strength of evidence may vary for different benefits and harms, making it difficult to rate the strength of evidence for the balance of benefits and harms.
2Data	Data may be lacking on the joint distribution of benefits and harms under various scenarios.
Scale of treatment effect	Benefits and harms may be reported on a relative or absolute scale. Systematic reviewers should conduct the quantitative assessment of benefits and harms on an absolute scale or use both absolute and relative scales.
Preferences	Values and preferences affect how people weigh the relative importance of outcomes. End users may perceive the incorporation of values and preferences in benefit harm assessments as the equivalent of making treatment recommendations.

RCTs = randomized controlled trials

Challenges With Regard to the Research Question

Populations

When assessing benefits and harms in systematic reviews, systematic reviewers face the challenge of determining whether the evidence on benefits and harms is applicable to the target population of interest. Systematic reviewers generally limit the inclusion of studies based on the characteristics of the study populations. The applicability of evidence to particular subgroups of interest (e.g., older or comorbid population) may be different for benefits than for harms. For example, the premarketing trials of the cyclo-oxygenase-2 inhibitors provided evidence of benefit in selected samples of patients with arthritis.[18] However, these trials excluded patients with comorbid cardiovascular disease. Subsequent studies demonstrated that cyclo-oxygenase-2 inhibitors increased the risk of serious adverse cardiovascular events, particularly in high-risk populations.[18]

Another challenge is that the ultimate balance of benefit and harm, as determined by decisionmakers, may vary for subgroups defined by specific characteristics among the population (e.g., age, or presence of comorbidity) or by estimated risk of adverse outcomes (e.g., <5, 5–10 or >10 percent risk for 10-year mortality). The reason for such variation may include the altered pharmacodynamic or pharmacokinetic properties of drugs with age. For example, a drug that is cleared by the kidneys and toxic at higher doses may accumulate in older populations and have a different benefit and harm profile than in younger populations.

Interventions

A systematic review may include studies with varying degrees of fidelity to the intervention. Fidelity is the extent to which patients in the study receive the same prestated intervention.[19] Lack of fidelity in the interventions may make it difficult for a systematic reviewer to interpret differences between studies in the reported benefits and harms. Measures of intervention fidelity include the adherence to the components delivered, exposure or dose of the intervention, quality of delivery, participant responsiveness, and uniqueness of the intervention.[20] The fidelity of the intervention to a prespecified protocol may vary between studies that report on benefit or harm.

Comparators

Another challenge is assessing whether the estimates of effect from active controlled trials reflect the beneficial effect of one intervention or the harmful effect of another. The efficacy of a pharmacologic intervention is often established in placebo-controlled randomized controlled trials (RCTs) that are required for regulatory approval.[12] The estimates of benefits and harms may vary based on the comparators used in RCTs, such as placebo or active controls. Estimates of treatment effect from active controlled trials may either be interpreted to reflect the beneficial effect of one intervention or the harmful effect of another.[18] In observational studies, comparisons of benefits and harms may also differ based on whether users of a therapeutic agent are compared with nonusers or users of other therapeutic agents. For example, when a five times higher risk of myocardial infarction was seen with rofecoxib 50 mg compared with naproxen 500 mg daily in the Vioxx® Gastrointestinal Outcomes Research study, one of the initial interpretations offered was that these findings reflected the beneficial effect of naproxen rather than the harmful effects of rofecoxib.[18] However, subsequent placebo controlled trials and other observational studies confirmed the cardiovascular hazards of rofecoxib and also provided evidence that naproxen offered no such cardiovascular benefit,[18] and may also carry a smaller cardiovascular hazard.

Outcomes

Systematic reviews usually consider a range of outcomes, with substantial variability and information asymmetry in the reporting of benefits and harms across studies. We generally see more reliable and more robust data on benefits as compared with harms of interventions. This may hinder an adequate assessment by decisionmakers of the balance of benefits and harms associated with the interventions of interest.[21]

Some systematic reviews report on composite outcomes. An assessment of the balance of benefits and harms using a composite outcome is challenging, especially if the individual elements of the composite outcome occur at different frequencies, show different effect sizes, or are of unequal clinical importance.[22]

Since there is no well-accepted measure of benefit across therapeutic areas, systematic reviews report on a range of surrogate and health outcomes.[21] Surrogate outcomes for benefit also pose challenges for assessing benefits and harms, and may include: biochemical endpoints, such as improvement in glycated hemoglobin or cholesterol; pathophysiological variables, such as improvement in blood pressure and ejection fraction; and morphological variables, such as left ventricular hypertrophy. Health outcomes are clinical outcomes that affect how patients feel, live or survive, such as quality of life, rate of survival, and patient satisfaction, and are sometimes referred to as patient-important outcomes. We use the terms "surrogate outcomes" and "health outcomes" (from the AHRQ Methods Guide) throughout this report. [23]

A recent systematic review found evidence of benefit from various oral hypoglycemic agents on the surrogate outcome of glycated hemoglobin.[12] However, the review found no conclusive evidence of the benefit of these agents on cardiovascular outcomes or mortality. To conduct an assessment of benefits and harms, reviewers must determine if the surrogate has been validated for each intervention for the health outcomes in the analytic framework.[23] Information may be unavailable to reviewers regarding links between surrogate and health outcomes and how these links may vary by subgroups such as older adults. Even if the surrogate is validated, it is challenging to translate the quantitative benefit of the surrogate outcome into a quantitative estimate for the health or patient-important outcome.

Studies generally prespecify and measure benefit outcomes with great reliability. This is not always true with harm outcomes, which are often unexpected and have much more variability in definition across RCTs. Unless researchers know the harms of an intervention *a priori* (e.g., bleeding with anticoagulants), researchers may not have a prespecified way of defining a particular type of adverse event or may include events in the category of "other adverse events." In a recent systematic review, the outcome of congestive heart failure reported as an adverse event ranged from heart failure events that are diagnosed only based on symptoms of breathlessness to congestive heart failure requiring hospitalization that is confirmed by echocardiography.[12] Systematic reviewers must assess the need for sensitivity, which increases the power to find evidence of rare harmful events, versus specificity of outcome definitions, which, in turn provides greater confidence in the strength of an association.

Time

The time horizon in studies may be inadequate for assessing all benefits and harms (e.g., early benefit, late risks), as the balance of benefit and harm in the short term may be different from the balance of benefit and harm in the long term. Although combined hormone replacement therapy did not prevent coronary events in the Heart and Estrogen/Progestin Replacement study, further analysis suggested early harm during the first year with the possibility of late benefit during followup 4 to 5 years later.[24]

Challenges With Regard to the Study Design

Well-designed RCTs provide the most valid estimates of the effect treatments have on benefits and harms, because of the control of confounding and selection bias. However, RCTs are often designed and powered to detect the effect of treatments on selected benefit outcomes, while harm outcomes receive less attention in terms of the quality of data ascertainment and statistical power. Also, some RCTs exclude patients that have certain characteristics (e.g., old age, comorbidity), associated with greater risk of harm. Some RCTs have a duration of followup that is sufficient for benefit outcomes but not for harm outcomes. In addition, RCTs may be

statistically underpowered for detection of rare but potentially serious harms because of the limited size and duration of trials. The resulting asymmetry in the quality and quantity of evidence for benefit and harm outcomes from RCTs are the main reason why some systematic reviews may include observational studies as well as RCTs.[25] Observational studies are an important source of information on benefit when RCTs are not long enough and/or do not include outcomes important to patients. Other nonrandomized designs, including spontaneous case reports, may provide a useful source of information on harms, particularly in the case of rare events with very low background rates in the general population. Estimates of harm from RCTs and observational studies vary due to differences in study quality, applicability (e.g., populations, interventions or comparisons), measurement of outcomes, publication bias, outcome reporting, or sources of funding.[25] As a result, estimates concerning harm tend to have more uncertainty compared with estimates of benefit.

Strength of Evidence

Systematic reviewers need to incorporate assessment of the strength of evidence when assessing the balance of benefits and harms in systematic reviews. The problem arises because systematic reviewers may grade the strength of evidence separately for each outcome of each key question.[4,9,12] For example, a recent systematic review on angiotensin converting enzyme inhibitors and angiotensin receptor blockers[4] presented evidence on various benefit and harm outcomes with different evidence grades.

The strength of evidence ratings for harm outcomes are more complicated than the ratings for benefit outcomes because definitions for harms are often not as explicit as they are for benefits. Harm assessments are also more complicated because of the need to incorporate lower-level evidence from heterogeneous data sources as well as the fact that studies on harms may not be as reliable, valid, or robust as the studies on benefits. These complicating issues create additional uncertainty in assessing the balance of benefits and harms.

No formal approaches exist that provide guidance on how to grade the strength (or quality) of the evidence on the balance, or comparison, of the benefits and harms of interventions. Thus it is difficult to incorporate the various grades of the strength of evidence on benefits and harms into a summary assessment of the balance of benefits and harms. Decisionmakers are likely to find it difficult to balance multiple outcomes reported on various scales (relative or absolute) with varying strength of evidence ratings. Although the Grading of Recommendations Assessment Development and Evaluation (GRADE) approach rates the overall quality of evidence by considering the lowest quality of evidence among critical outcomes,[26] that approach may not satisfy all decisionmakers.

Challenges With Regard to Available Data

Most systematic reviews are based on available summary (i.e., aggregate) data.[4,5] Thus, reviews usually only have access to marginal distributions of benefits and harms (i.e., separate results for each benefit and harm outcome) without information on the joint distribution of the outcomes (i.e., describing the correlation between benefit and harm outcomes). The joint distribution of the effects of treatment on the benefit and harm outcomes is seldom reported in studies. The studies only report separately on uncertainty, standard errors, and confidence intervals for each benefit and harm outcome. Without the joint distribution of all the effects, systematic reviewers have to assume independence of the benefits and harms, which may not yield a valid estimate of the uncertainty of the benefit and harm comparison metric. The joint

distribution can be obtained via bootstrap methodology if individual-level data are available from the studies. However, this is not possible in meta-analysis of summary data. Given the limitations of the data, systematic reviewers are challenged to report an estimate of uncertainty (e.g., 95% confidence interval) around the benefit and harm comparison estimate if and when they decide such a benefit and harm comparison estimate is useful for decisionmakers.

In rare cases, individual patient data are available to systematic reviewers.[3] Such data could include individual patient data from a subset of all studies, sufficient patient data from all studies, only marginally available patient data, or any combination thereof. The availability of individual participant data still requires careful consideration of salient analytical principles (such as whether all studies used an intention to treat analysis, or whether all studies had a sufficient duration of followup). For example, a recent individual patient data systematic review of inhaled budesonide among participants with chronic obstructive pulmonary disease reported no statistically significant increased risk of pneumonia with inhaled budesonide, but censored the analysis at 1 year of followup, despite having access to long-term data.[27] Followup analysis of the same dataset yielded effect estimates that could not rule out a clinically significant excess of pneumonia in long-term followup. In this example, it is unclear whether the summary data of the entire body of evidence is more reliable than individual patient data from a subset of the evidence.

Challenges With Regard to Scale of Treatment Effect

The analysis of benefits and harms in systematic reviews is challenged by the differences between the absolute or relative scales used for the analysis. The appropriate scale of analysis depends on assumptions about the causal effect. If the treatment is thought to have a multiplicative effect, then studies should use a relative risk scale to estimate the effect. If the treatment is thought to have an additive effect, then studies should use an absolute risk scale.[28] However, in many cases it is unclear which scale of analysis is most appropriate. Benefits could be gauged on an absolute, continuous, or relative scale. The relative scale is used most often for reporting treatment effects.[29] However, most quantitative approaches for assessing the balance of benefits and harms model data on an absolute scale, most likely because of the greater relevance of the absolute scale for clinical decisionmaking.[13,29] In a meta-analysis of harms, which are usually relatively rare events, reviewers may choose to model data on an odds or relative scale because of their strong statistical properties.

Challenges With Regard to Preferences

Systematic reviewers synthesize evidence on multiple benefits and harms to serve the needs of a variety of decisionmakers, such as guideline developers, payers, regulatory agencies, and ultimately patients and clinicians. Decisionmakers may have varying views about the relative desirability of different outcomes (i.e., preferences). For example, guidelines based on a systematic review of benefits and harms of strategies to prevent venous thromboembolism considered prevention of pulmonary embolism as a clinically important outcome, but prevention of deep vein thrombosis was not considered clinically important. [30] This was in contrast to other guidelines that considered both as clinically important outcomes.[31] The GRADE working group also provides some guidance on the importance of considering health state preferences when making a treatment recommendation.[26]

Systematic reviewers may elicit preferences about the relative importance of the different outcomes during various stages of the process by obtaining input from a variety of

decisionmakers or stakeholders. However, reviewers are challenged by how to make these preferences explicit. Although a systematic review that includes a key question on preferences related to outcomes and interventions (as opposed to the effectiveness of interventions) would be useful, such reviews are uncommon.

One problem is that systematic reviewers may find it difficult to incorporate values and preferences in benefit harm assessments without making treatment recommendations. Often, systematic reviewers are expected to review the evidence objectively without making specific treatment recommendations, leaving such decisions for policymakers that will have other contextual information (such as the availability of alternatives and the specific decisionmaking context). While including values and preferences regarding outcomes in a systematic review gets reviewers one step closer to making a treatment recommendation, it is still different from making treatment recommendations. Systematic reviewers should be aware of the narrow difference between drawing conclusions from a benefit harm assessment and making a treatment recommendation.

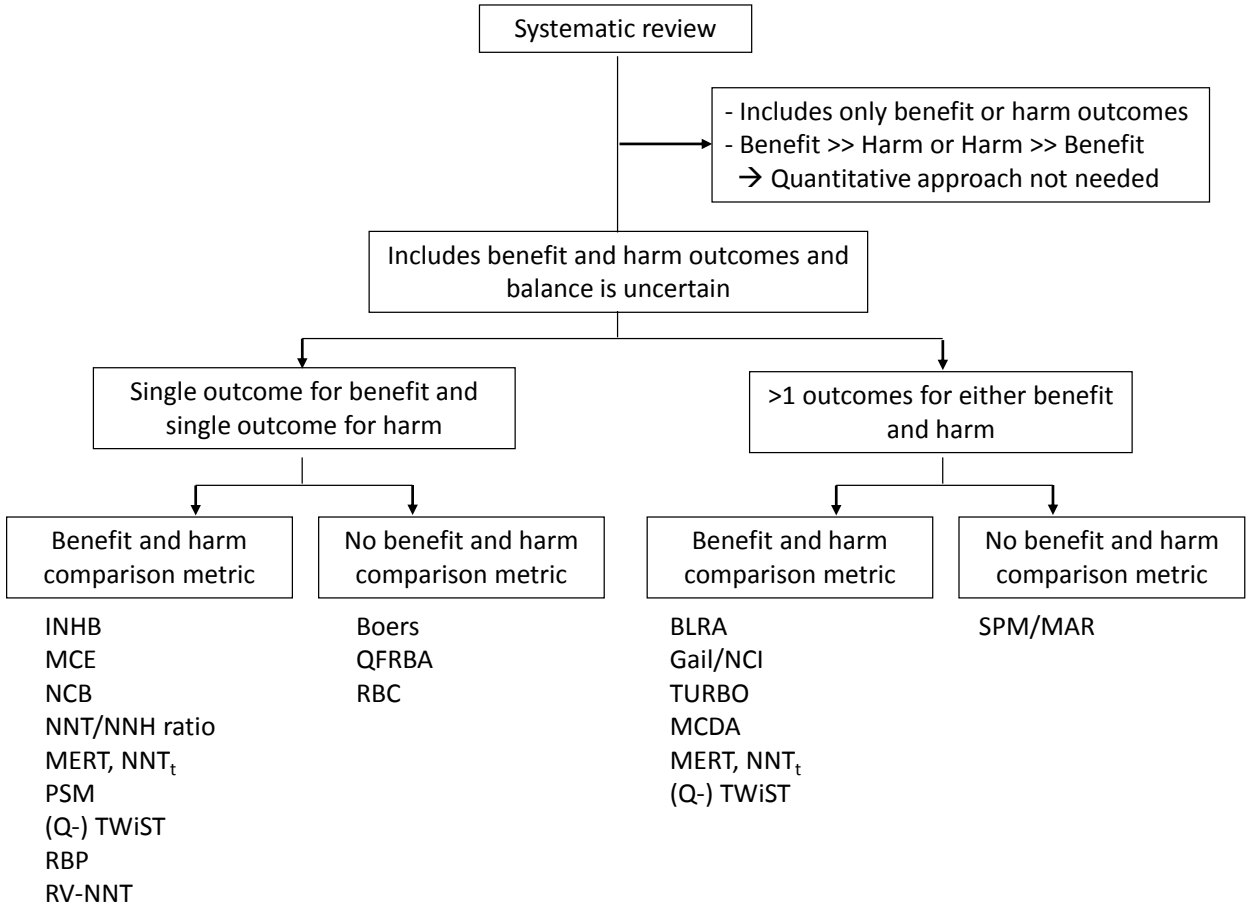
Methodological Characteristics of Existing Quantitative Approaches for Assessing Benefit and Harm

Past reviews of quantitative approaches for benefit and harm assessment have not organized the approaches according to important characteristics of the approaches.[13,16,17] Also, none of the reviews to date have considered the applicability of these methods for systematic reviews and meta-analysis. Systematic reviewers need a framework that recognizes the key characteristics of quantitative approaches, and organizes them accordingly. Such a framework could help to clarify the common and different elements of existing approaches, and guide further development of benefit harm assessments. A framework could also help to guide investigators and systematic reviewers when choosing an approach to benefit harm assessment. Most quantitative approaches are based on primary datasets, where investigators had control over study design, outcome selection, and individual patient data. We sought to review available quantitative approaches to assess the benefits and harms of medical and public health interventions, and to develop a framework for organizing the quantitative approaches to benefit harm assessment that would help systematic reviewers more effectively choose an appropriate method.

Literature Review and Classification of Quantitative Approaches for Benefit Harm Assessment

The team identified 16 quantitative approaches for benefit and harm assessment. Figure 2 (Framework for Organizing Quantitative Approaches) shows how we grouped these approaches into two broad categories. One category comprises simpler approaches that typically deal with one outcome for benefit (e.g., prevention of stroke) and one outcome for harm (e.g., gastrointestinal bleeding). These approaches can deal with composite outcomes that summarize multiple endpoints. Also, the approaches considering a single benefit and a single harm outcome can deal with several outcomes, but only in separate analyses.

Figure 2. Framework for organizing quantitative approaches



BLRA = benefit-less-risk analysis; Gail/NCI = Gail/National Cancer Institute approach: INHB = incremental net health benefit; MAR = maximum acceptable risk; MCE = minimum clinical efficacy; MCDA = multicriteria decision analysis; MERT = minimum target event risk for treatment; NCB = net clinical benefit; NNT = number needed to treat; NNH = number needed to harm; NNTt = threshold number needed to treat; PSM = probabilistic simulation methods ;Q-TWiST = (quality-adjusted) time without symptoms and toxicity; QFRBA = quantitative framework for risk and benefit assessment; RBC = risk–benefit contour; RBP = risk–benefit plane; RV-NNT = relative value adjusted number needed to treat; TURBO = transparent uniform risk benefit overview; SPM = stated preference method;

A second category includes more complex approaches that consider multiple benefits and harms in one analysis. However, some of these approaches can be used for both single and multiple outcomes.

In Figure 2, we categorized these approaches according to how the medical literature typically uses them (e.g., number needed to treat [NNT] and number needed to harm [NNH] are typically used for single outcomes). We listed a few of them in two categories if we could not clearly categorize them (e.g., minimum target event risk for treatment, and quality-adjusted time without symptoms or toxicity [Q-TWiST]). Figure 2 also demonstrates that a benefit and harm comparison metric that puts all outcomes on a common scale further distinguishes between approaches.

Some systematic reviews might not need quantitative approaches at all. We can think of two such situations: (1) systematic reviews that exclusively focus on either benefits or harms; or (2) treatments where benefits are much greater than harms or vice versa). This situation is relatively rare. Otherwise, systematic reviewers need to compare carefully the results for benefit and harm outcomes before deciding against a quantitative benefit harm assessment.

One example is exercise therapy[32] for patients with chronic obstructive pulmonary disease. Exercise therapy provides great benefit to patients (e.g., improvement of symptoms, exercise capacity, and quality of life) with very little harm (e.g., small risk of accidents during unsupervised exercise) or inconvenience (e.g., going to a rehabilitation center). In such situations, the decisionmaking is very unlikely to change if the results from a quantitative benefit harm assessment are available. However, systematic reviewers should not make premature judgments about the need for a quantitative benefit harm assessment based on their own preferences for or against certain interventions.

For a particular systematic review, not all 16 approaches are a sensible option. Usually, the best approach depends on the number of outcomes, the need for a benefit and harm comparison metric, and the quality and quantity of available data Table 2).

Table 2. Organizing framework of existing quantitative approaches for benefit harm assessment

Key characteristics	BLRA	Boers	Gail	INHB	MCDA	MCE	NCB	NNT& NNH	PSM	QFRBA	Q-TWiST	RBC	RBP	RV-NNT	SPM & MAR	TURBO	Number of approaches having each characteristic
Types of data																	
Require individual patient data	Yes	Yes	No	No	No	No	No	No	No	No	No	No	No	No	Yes	No	Yes: 3 No: 13
Types of analyses																	
Data driven versus simulation	DD	DD	DD/S	DD	DD	DD	DD/S	DD	S	DD	DD	DD/S	DD	DD	DD	DD	DD: 15 S: 4
Types of B&H metrics																	
Absolute versus relative measures versus QALY versus other	Other	A	A	QALY	A / R	A	A	A	A	A / R	A /QALY	A	A	A	A	A / R	A: 14 R: 3 QALY: 2
Assumptions																	
Put B&H outcomes on same scale	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	Yes	No	Yes	Yes: 12 No: 4
Uncertainty estimates for B&H assessment	No	No	No	No	Yes	No	Yes	No *	Yes	NA	No	Yes	No	No	NA	No	Yes: 4 No: 10 NA: 2
Joint distribution of B&H outcomes considered for uncertainty estimates	No	NA	P	NA	No	NA	No	No	P	NA	NA	P	P	No	NA	NA	P: 4 No: 5 NA: 7
Multiple endpoints versus composite outcomes for B&H	M	Comp	M	M	M	M	M	M/ Comp	M	M	M	M	Comp	M	M	Comp	M: 13 Comp: 4
Consideration of preferences																	
Explicitly considers preferences for B&H assessment:	Yes	No	Yes	No	Yes	Yes	Yes	No	Yes	No	Yes	No	No	Yes	Yes	No	Yes: 9 No: 7
Types of presenting benefit risk balance																	
B&H difference versus B&H ratio versus Time gained/lost vs. B&H graphic versus other	D	Graph	D	D	D, Ratio	D	D	Ratio	D	D, Ratio	Time, D	Graph	Graph	Ratio	D	Graph	D: 10 Ratio: 4 Graph: 4 Time: 1

*For some variants of the NNT approach such as NNTt and the minimum target event risk for treatment (MERT), uncertainty estimates exist[53]

A = absolute risk metric; B&H = benefit harm assessment; BLRA = benefit-less-risk analysis; D = difference; DD = Data Driven; Graph = graphic; INHB = incremental net health benefit; O = other; M = multiple; MAR = maximum acceptable risk; MCDA = multicriteria decision analysis: MCE = minimum clinical efficacy; NA = not applicable; NCB = net clinical benefit; NNT = number needed to treat; NNH = number needed to harm; P = possible; PSM = probabilistic simulation methods; QFRBA = quantitative framework for risk and benefit assessment; Q-Twist = (quality-adjusted) time without symptoms and toxicity; R = relative risk metric; RBC = risk–benefit contour; RBP = risk–benefit plane; RV-NNT = relative value adjusted number-needed- to-treat; S = simulation; SPM = stated preference method; TURBO = transparent uniform risk benefit overview

To facilitate the comparison of these approaches in the context of a systematic review, the team selected, as examples, four approaches to give a more detailed description. The NNT/NNH ratio is an example of an approach where a single benefit outcome and a single harm outcome are of interest. The Multicriteria Decision Analysis considers multiple outcomes. The Gail/National Cancer Institute (Gail/NCI) method places multiple outcomes on a single benefit and harm comparison metric. Finally, the risk-benefit contour is a graphical approach that visualizes the probability of harm and benefit and associated uncertainty. A more extensive evaluation of some of these approaches will be presented in our next report.

Number Needed To Treat and Number Needed To Harm

The number needed to treat (NNT) and its harm counterpart, the number needed to harm (NNH), are perhaps the most widely used measures of benefit and harm when presented separately in systematic reviews. Also, practice guidelines most commonly use NNT when discussing benefit and harm balance. NNT and NNH refer to the number of individuals that need to be treated over a specified period of time for one person to benefit or be harmed, respectively, and will therefore vary as the specified treatment time varies.

The NNT and NNH are almost always presented separately (i.e., not using a benefit and harm comparison metric such as the ratio of number-needed-to treat [NNT] to number-needed-to harm [NNH], hereafter referred to as the NNT/NNH ratio). For example, the Clinical Practice Guidelines on Antithrombotic Therapy in Atrial Fibrillation of the American College of Chest Physicians present NNTs based on a systematic review of RCTs of oral anticoagulant therapy compared with no antithrombotic therapy: The efficacy of warfarin was consistent across studies with an overall relative risk reduction of 68 percent (95% confidence interval, 50 to 79 percent) analyzed by intention-to-treat.[33] The absolute risk reduction implies that 32 ischemic strokes will be prevented each year for every 1,000 patients treated (or 32 patients needed to treat for 1 year to prevent one stroke, NNT = 32).

In contrast, studies do not commonly use the NNT/NNH ratio. One reason for the rare application of this benefit and harm comparison metric may be that investigators or guideline developers are reluctant to implicitly weigh benefit and harm outcomes equally on the same scale because of uncertainty about their relative importance.

Multicriteria Decision Analysis

The analytic hierarchy process (AHP) is an example of a multicriteria decision analysis approach. We use the AHP to explain the principle of a multicriteria decision analysis using the systematic review of oral hypoglycemic agents for type 2 diabetes. The first step in AHP analysis consists of defining the goal of the decision, the alternatives being considered, and the criteria that determine how well the alternatives can be expected to meet the goal.[34,35] Studies organize these into a hierarchical decision model with the goal of determining the best treatment for type 2 diabetes.

Operationally, we could define two criteria as being necessary for determining the best treatment: (1) it maximizes benefits via glucose reduction, and (2) it minimizes harms or medication related adverse effects. We could divide the criteria on maximizing benefits into three sub-criteria: health-related quality of life, microvascular benefit (such as improvements in incidence of neuropathy, nephropathy, and diabetic retinopathy), and potential macrovascular benefit. We could subdivide the criteria on minimizing risk into six sub-criteria of medication-

related adverse events: congestive heart failure, fractures in women, macular edema, bladder cancer, myocardial infarction, and hypoglycemia.

In the second step, reviewers obtain information about how well the alternatives can be expected to fulfill the decision criteria from the systematic review. The third step consists of two parts: (1) comparing the ability of the alternative treatments to fulfill the prespecified criteria (maximizes benefit and minimizes harm), using standard AHP pair-wise comparisons, and (2) assessing the importance of these criteria to the decision goal. In the fourth step, reviewers take the scales created in step three and combine them to create a summary score indicating how well they can expect alternative treatments will meet the decision goal. The fifth step consists of sensitivity analyses to explore the effects of changing the estimates or judgments used in the original analysis.

The main advantages of AHP are: the use of the summary score; the incorporation of uncertainty; and the option to explore the extent to which every criterion, judgment, and weight contributes to that score.

The Gail/National Cancer Institute

Some decisionmaking contexts are more complicated because of the many different treatment outcomes as well as many sources of uncertainty they include. A well-known example is the use of tamoxifen for the prevention of breast cancer. Tamoxifen reduces the risk for invasive and in situ breast cancer substantially, and it also prevents some bone fractures.[36] However, it also increases the risk for endometrial cancer, stroke, and pulmonary embolism.

The National Cancer Institute (NCI), under the leadership of Dr. Mitchell Gail, developed an approach for dealing with multiple outcomes. Rather than simplifying the benefit harm assessment to single outcomes, they estimated the probability of various outcomes for women with and without tamoxifen therapy over a period of 5 years. Based on observational studies, surveillance registries, and placebo arms of RCTs, they first estimated the expected number of invasive breast cancers, in situ breast cancers, hip fractures, endometrial cancers, strokes, pulmonary emboli, deep vein thromboses, Colles' fractures, spine fractures, and cataracts in the absence of tamoxifen treatment (each per 10,000 women and over 5 years). They estimated these numbers overall and also stratified for different age and race categories.

They then estimated, for each outcome and based on the Breast Cancer Prevention Trial, the expected number of the same outcomes with tamoxifen treatment (each per 10,000 women and over 5 years).[14] Here, they also estimated the numbers overall and stratified for different age and race categories. To put all outcomes on the same scale, but to also consider the relative importance of these outcomes, they categorized the outcomes into life threatening, severe, and other outcomes; and suggested weighting them with some factor (e.g., 1 for life threatening, 0.5 for severe, and 0 for other outcomes). These categories and weights could be modified according to patient or treatment-provider preferences. Sometimes, it may be difficult to choose the weights, because of a lack of evidence on patient preferences. In such cases, sensitivity analyses should take different weights into consideration.

Ultimately, researchers present the results of the benefit harm assessment as the net number of events prevented per 10,000 women treated with tamoxifen over a period of 5 years. For example, for a 45-year-old woman with her uterus and a 4 percent risk of invasive breast cancer over 5 years, the net number of events prevented (weighted by clinical importance) was 196 per 10,000 women with this profile. The expected number of prevented invasive and in situ breast cancers was 299 per 10,000 women, with 59 per 10,000 women having some harm such as

endometrial cancer, stroke, pulmonary embolism, or deep vein thrombosis. The net benefit (benefit minus harm events) varied considerably and was positive for some profiles but negative for others (e.g., black women age 50-59 years and a 5-year risk of invasive breast cancer of 4 percent).

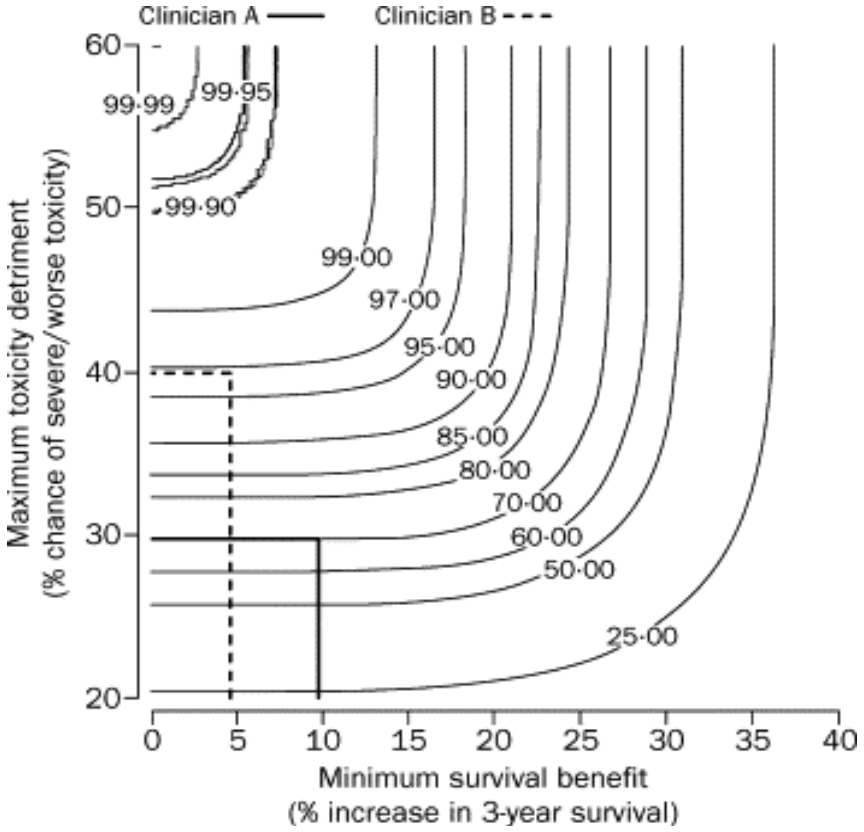
Risk-Benefit Contour

The risk–benefit contour plot is a graphical method used to visualize the probability of benefit and harm and associated uncertainty. It portrays the probability of benefit for a new treatment compared with another treatment against the probability of harm for that new treatment as compared with another treatment.[37]

The probabilities that tell us how likely it is that a treatment is beneficial can be derived from standard statistical software or a Z table of normal values. For example, the 95% CI of a (hypothetical) relative risk of 0.75 may range from 0.50 to 1.00, which tells us that we can have 5 percent confidence that the effect is outside this range (half of which [2.5 percent] is above 1.00). Thus, for each effect estimate, the probabilities of an effect can be calculated based on the published reports of RCTs (i.e., point estimates and 95% CIs).

Contour lines portray the shape of this relationship for a number of different probabilities and confidence levels. As shown in Figure 3, a Risk-benefit Contour Example, a clinician might recommend the new treatment if there is at least a 10 percent survival benefit compared with another treatment, and if the probability of severe harm is not increased by more than 30 percent, compared with the other treatment. The contour lines show a 70 percent probability that the new treatment provides 10 percent survival benefit, and not more than 30 percent chance of severe harm, compared with the treatment alternative.

Figure 3. Risk-benefit contour example



Adapted from: Fig. 6, in Shakespeare et al. Lancet 2001 [37]

The clinician and patient may or may not accept the uncertainty associated with a 70 percent probability when deciding for or against the new treatment. The risk-benefit contour plot is a way to express uncertainty associated with certain pairs of benefit and harm.

Description of Key Characteristics of 16 Quantitative Approaches for Benefit Harm Assessment

Quantitative approaches differ in more than just the number of outcomes considered and the use of a benefit and harm comparison metric. Therefore, we focus here on additional characteristics of benefit harm assessment approaches that systematic reviewers should consider when choosing the approach for a specific key question and context. Previous reviews have discussed some of these methods, but not in the context of systematic reviews.[13]

For the purpose of this report, we identified the following six key characteristics of quantitative approaches for benefit harm assessment:

(1) The type of data needed: Individual patient data have advantages over the aggregate data typically available for evidence synthesis. An individual study should have information on the occurrence of benefit and harm outcomes for each patient as well as on their temporal sequence. For each patient, data can describe the cumulative occurrence of benefit and harm outcomes over time as well as their correlation. For example, phosphodiesterase-4 inhibitors are new drugs for patients with chronic obstructive pulmonary disease that aim at reducing exacerbations, but they

also have some gastrointestinal toxicity. Patients who take these drugs regularly have a lower risk for exacerbations than those who do not take them, but they are also at greater risk for nausea, diarrhea, and abdominal pain.[38] A patient with greater susceptibility to gastrointestinal toxicity is more likely to have more frequent gastrointestinal symptoms over the course of treatment. Patients on the lower dose and those who are less susceptible experience a lower frequency of adverse effects. An important aspect of the availability of such individual data is that researchers can consider the joint probability of benefit and harm outcomes.

(2) The type of analyses: Analyses for benefit harm assessment could include any type of statistical analysis. We can make a major distinction, however, between approaches that are data driven (deterministic) and approaches that use modeling (e.g., stochastic), where data are not described by unique values, but rather by probability distributions.

(3) The type of benefit and harm comparison metric: Studies can use comparison metrics based on absolute differences or relative differences. Alternatively, studies could use quality-adjusted life-years as a benefit and harm comparison metric. Each metric has its advantages and disadvantages and the choice depends on the decisionmaking context.

(4) Assumptions: In quantitative approaches to benefit harm assessments, researchers usually make assumptions. For example, if they use a benefit and harm comparison metric, the assumption is that the outcomes considered are the ones that inform decisionmaking and that it makes sense to combine the outcomes on a single scale. It may not make sense to combine a surrogate outcome (e.g., lung function) and quality of life on the same scale. Outcomes may not be of equal importance and can be weighted according to the importance to patients. Other assumptions relate to the joint occurrence of separate outcomes. Some approaches assume that separate outcomes occur independently, which may or may not be justifiable.

(5) Consideration and incorporation of patient preferences: Some quantitative approaches explicitly consider patient preferences for different outcomes when weighing the benefits and harms. For example, aspirin may prevent 10 major strokes in 1,000 middle-aged men treated for 10 years, as compared with no aspirin prevention.[39] Aspirin may also cause an additional 40 major gastrointestinal bleedings over the same prevention period, as compared with no aspirin prevention. If patient preferences are equal for major stroke and major gastrointestinal bleeding, it would not be advisable to use aspirin in middle-aged men. However, if patients weigh major stroke as being 10 times as important as major gastrointestinal bleeding, aspirin may provide more benefit than harm.

(6) Formats for presenting benefit and harm balance: Studies can use various formats to present the results of the quantitative benefit harm assessment. The studies can present the benefit and harm balance as a difference in the number of events between a treatment options, or they can describe it as a ratio. The studies might also express the balance by the time gained or lost without symptoms through a treatment, or use graphics that depict the benefit and harm balance for patients at different outcome risks or based on other patient characteristics.

Brief Description of Key Characteristics of Each Approach

Here we first summarize the 16 approaches according to key characteristics and then provide a brief discussion of each approach. Three approaches require individual patient data (See Table 2 for a summary of all of the approaches), and 13 approaches do not require individual patient data. Fifteen of the approaches are data driven, but 3 of the 16 may also use simulation while one approach, probabilistic simulation, is based entirely on simulation. Twelve of the 16 approaches put benefit and harm outcomes on the same scale to provide a benefit and harm comparison

metric. Only 4 approaches provide measures of uncertainty around the benefit and harm comparison metric, although considering uncertainty is likely to be of importance for decisionmakers and organizations making treatment recommendations. Four approaches could consider the joint distribution of benefit and harm outcomes for the estimation of uncertainty, but many examples using these 4 approaches do not consider the dependence between benefit and harm, and only consider their marginal distributions (i.e., they consider them to be independent). Four approaches use composite outcomes for benefit and composite outcomes for harm; while 13 approaches use multiple outcomes (NNT and NNH can use both). Nine of the 16 approaches explicitly incorporate patient preferences.

The team found that most of the approaches could be used in systematic reviews since most used aggregate data. This important finding means that systematic reviewers have a wide range of quantitative approaches for benefit and harm assessment and can choose the most appropriate and feasible approach for a specific question.

Benefit-Less-Risk Analysis

Benefit-less-risk analysis (BLRA) combines benefit and harm into a single metric, and was designed primarily for clinical trials.[40] BLRA takes advantage of individual patient data. For each patient of a trial (who is under some or no treatment), researchers record a benefit as yes = 1 or no = 0, and express the harm as a value between 0 and 1.

This type of analysis presents the relationship between benefit and risk as risk subtracted from benefit (e.g., 1 for benefit – 0.2 for harm). BLRA thus allows for statistical testing of comparisons between treatment groups and can consider patient preferences, expressing the relative importance of benefit and harm outcomes. If a systematic review used this method, the review would need to gather individual patient data from the primary studies.

Boers' 3x3 Table

This quantitative approach does not require any statistical models but offers a way for organizing outcome data on the same scale.[41] This approach needs individual patient data. Researchers split the outcomes of patients into three categories (minimal, moderate, or major benefit, and minimal, moderate, or major harm) and display the number of patients with a certain benefit harm profile (e.g., major benefit and minimal harm) in a 3x3 table. The approach does not consider treatment effects directly, since studies would need a separate 3x3 table for each treatment group. As a consequence, no measures of uncertainty are available. It does not consider patient preferences, but includes instead the clinicians' view or agreement as to what constitutes minimal, moderate, or major benefit or harm, respectively.

The method is feasible for both single trials and systematic reviews. A disadvantage is that, although each table is simple and easy to read, it requires readers to somehow estimate treatment effects across tables or to provide a benefit and harm comparison metric. Thus, the method challenges rather than facilitates conclusions concerning benefit and harm.

The Gail/National Cancer Institute

This approach is probably one of the most comprehensive approaches for benefit and harm assessment, since it considers various data sources to balance the benefits and harms of a treatment. As described earlier in this report, the Gail/NCI approach calculates a benefit and harm comparison metric as the sum of benefit and harm outcome rates per patient profile. It

incorporates patient preferences by looking only at one severity grade or by weighting outcome rates that reflect patient perception of very severe, severe, or moderately severe events.

It does not provide estimates of uncertainty arising from sampling variation (although additional analysis such as probabilistic simulation can provide measures of uncertainty) or from combining different data sources of different methodological quality. However, by looking at benefit and harm comparison estimates across patient profiles, one gets an impression of how the net benefit changes, even qualitatively, as the baseline harm changes.

This approach is resource-intensive because it considers multiple data sources and multiple outcomes. However, a comparative effectiveness review could provide an ideal basis for this approach if reviewers collected additional data, such as risk for outcome estimates from observational studies. The U.S. Preventive Services Taskforce used a similar, but simplified, approach to make recommendations on the use of aspirin for the prevention of myocardial infarction. Similar to the tamoxifen example, it estimated the number of benefits (myocardial infarction avoided) and harms (bleedings) per 1,000 men or women based on observational data, and it combined the evidence on treatment benefit and harm with these outcome estimates. The benefit and harm comparison metric provided the number of net events (benefit minus harm) prevented or in excess when aspirin was used.[39]

Incremental Net Health Benefit

Incremental net health benefit provides a benefit and harm comparison metric (using quality-adjusted life years [QALYs]) to place one or more benefit and harm outcomes on the same scale, and it calculates the difference between benefits and harm between treatments (thus a result greater than 0 is favorable).[42,43] A requirement for this approach is either the valid measurement of utilities or the transformation of quality-of-life scores into utilities (which is sometimes inaccurate). Also, it may be difficult to disentangle benefits and harms when using utilities because the utility for any given health state could be based on a combination of benefits and harms.

Multicriteria Decision Analysis

Multicriteria decision analysis allows for systematic decisionmaking in complex situations involving tradeoffs, by considering various benefits and harms associated with treatments.[44,45] This analysis uses a decision-tree model to incorporate benefits from clinical trials and harms such as adverse effects. It allows for input from various stakeholders who may assign different preference weights to the risks and benefits. Multicriteria decision analysis represents an approach to reducing the multidimensionality of benefit harm assessment in a systematic way, and it makes judgments both explicit and transparent.[46] It allows for decisionmaking in the presence of uncertainty, and it can incorporate data from multiple sources including systematic reviews.[47]

The challenges of its application to systematic reviews include getting reliable information on various preferences, agreement on all relevant benefits and harms (and the relative importance and weighting of these outcomes), and the need to specify a decisionmaking context given that systematic reviews are usually conducted to meet the needs of multiple decisionmakers. The flexibility of multicriteria decision analysis also poses challenges for benefit harm assessment since systematic reviews often are unable to provide sufficient evidence on all relevant inputs, especially less tangible inputs (e.g., societal values, opportunity costs) that may alter the harm-benefit balance in a particular decisionmaking context.

Minimum Clinical Efficacy

Minimum clinical efficacy assesses benefit by comparing benefits and harms on a probability scale, where it applies relative risk reductions (treatment benefits) and risk increases (harms) to absolute probabilities as observed in untreated groups.[46,48] It rates an intervention as having at least minimal clinically efficacy if the difference between benefit and harm is positive or above a minimally acceptable threshold. Minimum clinical efficacy can consider relative utilities, but a limitation includes the inability to provide uncertainty estimates for the benefit and harm comparison metric.

Net Clinical Benefit

Similar to the Gail/NCI approach, the calculation of the net clinical benefit considers different data sources such as RCTs, observational studies, and patient preferences, and it provides profile-specific benefit and harm comparison estimates.[49] Net clinical benefit calculates the benefit and harm comparison metric as the sum of all expected benefits minus the sum of all expected harms. It calculates the benefit from the pooled relative risk reductions (based on meta-analysis) that are applied to patients at different risk for the benefit outcome (e.g., stroke). It calculates the expected harm from the risks for the harm outcome and the patient preferences for the harm outcome. It calculates net clinical benefit using a Bayesian approach where all steps (meta-analysis, calculation of expected benefit, and expected harm) are modeled simultaneously.

A major advantage of this approach is its flexibility with regard to combining different data sources and placing distributions on each parameter. Thereby, researchers can quantify uncertainty around the parameters. Net clinical benefit considers patient preferences for different outcomes, but similar to other approaches, the selection of particular values for preferences has a large impact on the net clinical benefit estimates. Figure 2 categorizes the approach as an approach that considers only single benefit and harm outcomes, because published applications of the approach considered only one benefit and one harm outcome. The approach offers, theoretically, enough flexibility to consider multiple outcomes.

Number Needed To Treat and Number Needed To Harm

NNT and the NNH are the number of individuals who need to be treated over a specified period of time for one person to experience the benefit or the harm, respectively.[50,51] NNT and NNH depend on baseline risk (and are thus sensitive to different patient profiles) and the degree of relative risk reduction provided by the intervention, which is often assumed to be constant across the disease spectrum but may actually vary.

Studies cannot calculate NNT and NNH for continuous outcomes unless such outcomes are dichotomized. Systematic reviews often use NNT and NNH as a measure of benefit and harm. Studies can calculate NNT and NNH for single outcomes (e.g., NNT for exacerbations vs. NNH for fractures) or for composite outcomes for both benefit and harm.

An advantage of this method is that it usually keeps benefit and harm separate, and it leaves room for incorporation of preferences by decisionmakers and consideration of multiple outcomes. Studies can calculate NNT to NNH ratios (NNT/NNH ratios) or NNT to NNH differences. Since the concept of NNT is one of frequency and not of importance, studies should only calculate NNT/NNH ratios or NNT to NNH differences for outcomes of similar importance, unless they are weighted.[52] When studies calculate an NNT/NNH ratio or NNT to NNH

differences as a benefit and harm comparison metric, researchers assume their independence and may need to extrapolate the ratio or difference so that they refer to the same time period.

Extensions of the NNT/NNH ratio approach are the threshold NNT (NNTt), the minimum target event risk for treatment, and the subject-year adjusted NNT.[53] The NNTt reflects the point at which the risks and costs of a clinical intervention balance the benefit, and the minimum target-event risk for treatment defines the minimum target-event risk at which the intervention is justified. Subject-year adjusted NNT refers to the denominator as being subject-years instead of participants, to better account for time on treatment for participants. For example, if there are two events per 1,000 subject-years in the control group and one event per 1,000 subject-years in the intervention group, the NNT is 1,000 subject-years, which means that, with treatment, one fewer event would occur with every 1,000 subject-years.

Methods for providing uncertainty for these benefit and harm comparison metrics are available.[53] The NNT/NNH ratio, NNTt, and minimum target-event risk for treatment all are feasible within a systematic review context.

Probabilistic Simulation Methods

The probabilistic simulation method employs probabilistic simulations for benefit and harm comparison estimates using Monte Carlo methods. The probabilistic simulation method estimates the incremental benefit versus the incremental harm for only one benefit and one harm outcome in a single model. Multiple outcomes require different models (similar to the NNT/NNH ratio).

This method can incorporate parameters from multiple data sources (e.g., systematic reviews of RCTs and observational studies), patient preferences (e.g., from conjoint analysis), and different patient profiles.[54-56] It estimates uncertainty around the benefit and harm comparison estimate, with or without consideration of the joint distribution of benefit and harm (depending on the availability of individual-level data or reporting of covariance). Probabilistic simulation methods therefore may provide a comprehensive approach to assessing benefit and harm.

Quantitative Framework for Risk and Benefit Assessment

A quantitative framework for risk and benefit assessment reports on benefit and harm separately. It does not provide a benefit and harm comparison metric and uncertainty estimates are only available for the separate treatment effects for benefit or harm outcomes.[13]

An advantage of this method is that keeps benefit and harm separate, and it leaves room for incorporation of preferences by decisionmakers and consideration of multiple outcomes. Also, a quantitative framework for benefit and harm assessment is probably the way most meta-analyses currently report or discuss the benefit and harm assessment.

(Quality-Adjusted) Time Without Symptoms and Toxicity

Time without symptoms and toxicity (TWiST) compares treatments in terms of the time gained without symptoms versus the time lost due to the experience of adverse effects.[57,58] It puts the benefit and harm on the same scale (e.g., time). Q-TWiST is a further development that converts time into QALYs.[59] Here, the benefit and harm comparison metric is the difference between the treatment-associated gain in QALYs and the loss in QALYs associated with adverse effects of treatment. Oncology has widely used Q-TWiST.

The major advantage of this method is the ability to incorporate patient preferences, which may change over time. The method depends heavily on the availability of measurements that estimate the length of time without symptoms that estimate the time during which adverse effects were experienced, and that have the ability to distinguish benefit and harm. For example, quality of life and some preference-based instruments often provide a composite score that already synthesizes the overall experience of a patient. QALYs value health states rather than changes in health states, and lack of a measure of uncertainty around these measurements may limit the usefulness of Q-TWiST. This method may be difficult to apply in a systematic review, since primary studies are unlikely to report QALYs associated with benefits and harms.

Risk-Benefit Contour

The risk–benefit contour plot is a graphical method for assessing benefits and harms. It portrays the probability of benefit for a new treatment compared with another treatment against the probability of harm for that new treatment (as compared with another treatment).[37] Contour lines portray the shape of this relationship for a number of different probabilities and confidence levels. The risk–benefit contour plot is a way to express uncertainty associated with certain pairs of benefit and harm. The plot conveys study-level relationships, and it does not consider the interdependence of the probability of benefit and harm at the individual level.

Although the method does not incorporate weights (representing patient preferences) for each type of outcome, researchers could adapt it to do so. Researchers should probably view risk–benefit contour plot as a way to present data and visualize uncertainty, whereas they can base the underlying analyses that yield the probability estimates on different statistical approaches such as various forms of probabilistic simulation methods.

Risk-Benefit Plane and Risk-Benefit Acceptability Threshold

Risk–benefit plane, also known as risk–benefit acceptability threshold, displays both separate estimates of benefit and harm and a benefit and harm comparison metric in a simple figure.[56] It does not consider the individual level interdependence between benefit and harm.

Using an absolute scale, the probability of benefit (from a comparison between two treatments) is plotted against the probability of harm. Studies call the slope created by a line between the origin and the two-dimensional result the risk-benefit acceptability threshold. This method does not consider outcome weights that would reflect patient preferences.

Relative Value Adjusted Number Needed To Treat

The major advantage of relative value adjusted NNT over NNT and NNH is that it allows for incorporation of preferences into assessing benefit and harm.[46,48] Otherwise, it offers the same advantages as the NNT/NNH ratio approach, and suffers from the same limitations. Systematic reviews would need information on preferences to use this method.

Stated Preference Method or Maximum Acceptable Risk

Stated preference methods elicit patient preferences for various tradeoffs and their acceptability for treatment. Studies use the stated preference method to survey patients as to the amount of burden from adverse effects they are willing to accept to experience the benefit from treatment (Maximum Acceptable Risk).[60-64] Researchers need individual patient data on preferences for these approaches.

The typical method to elicit preferences is discrete choice or conjoint analysis, where respondents have to pick their preferred treatment from two treatment scenarios that characterize the benefit and harm of these treatments. These approaches assume that the attractiveness of a particular treatment is a function of the benefit and harm attributes, which are combined in various ways in different vignettes of the survey.[65]

Transparent Uniform Risk-Benefit Overview

The transparent uniform risk-benefit overview (TURBO) diagram displays the factors "R" and "B". "R" is the sum of the most serious adverse effect (scored from 1–5) and the second most serious adverse effect (scored from 1–2).[15] This approach is based on the frequency and severity of the harm outcome. Similarly, "B" is the sum of the primary benefit (scored from 1–5) and the ancillary benefit (scored from 1–2). The approach bases this score on the probability and extent of the benefit outcome. The "T" score represents the benefit and harm comparison metric and ranges from 1 (high "R" and low "B" score) to 7 (high "B" and low "R" score).

Studies typically use the TURBO in a regulatory context (e.g., the European Medicines Agency) and therefore they base it on single trials, but it can easily be extended to systematic reviews. The factors "R" and "B" can be based on absolute or relative measures of treatment effects for which uncertainty estimates are available. However, the "T" score has no uncertainty estimates.

Unlike other approaches, the TURBO explicitly considers not only one, but two outcomes for both benefit and harm that are weighted differently. Challenges to applying the TURBO method include arbitrary selection of the two benefit and harm outcomes from a comprehensive list of outcomes and the way scores (combining frequency and importance of outcomes) are assigned.

Desired Properties of Quantitative Benefit and Harm Assessment

Based on our review of the characteristics of existing quantitative approaches to benefit and harm assessment, we identified desired properties of quantitative benefit and harm assessments.

Study Population

A quantitative assessment of benefits and harms should be derived from a study population that covers the range of subjects for which a quantitative benefit and harm assessment is relevant. It should consider the health care setting in which particular decisions are taken, indicators of disease severity, socio-demographics, and comorbidity (e.g., those with two or more chronic conditions). The study population does not need to be particularly broad, but the key is that the study population used in the benefit and harm assessment reflects the target population.

Selection of Outcomes

Ideally, a benefit and harm assessment should include all benefit and harm outcomes that could influence decisionmaking. The selection of outcomes depends on the decisionmakers and those affected by the decisions, which could be patients, health care providers, policymakers, or payers. For many treatment decisions, multiple benefit and harm outcomes will be necessary.

Quality of Outcome Measurement

Another desired property of a quantitative benefit and harm assessment is the availability of high-quality data for both benefit and harm outcomes. As mentioned previously, RCTs are commonly designed to provide high-quality evidence and are usually powered to detect statistically significant differences on efficacy or benefit outcomes, but harms often receive less attention in terms of accurate and valid measurement. Therefore, observational studies should also be considered for harm outcomes.

Outcome Assessment Across Studies

Ascertaining benefit and harm outcomes should be accurate and similar across studies (e.g., how a cardiovascular event, an exacerbation, or pain is defined). Whether this is the case depends much on the disease area. While in some areas benefit and harm outcome measurement is harmonized (e.g., outcome measures in rheumatology)[66], the selection of outcomes and their measurement varies widely in other disease areas.

Study Duration

Any quantitative benefit and harm assessment should include patient outcomes for the entire period of treatment exposure and followup. This means the assessment should include studies that follow patients with an outcome (benefit or harm) until death (if possible). This is a commonly occurring problem especially when the followup is stopped after a benefit-related outcome, since this could result in an underestimation of harms of the intervention.

Evidence Selection

A quantitative benefit and harm assessment should be based on the best evidence available, and the evidence should be comprehensive. A systematic review should underlie any quantitative benefit harm assessments. In this way, reviewers identify all studies that potentially contribute to the quantitative benefit harm assessment. The quality of the evidence that reviewers consider will depend on the underlying study designs and on the desired properties for the quantitative benefit harm assessment. As a consequence, a tradeoff between experimental and observational data is often necessary. While trials provide a higher quality of treatment effects (for both benefit and harm outcomes), observational studies are often the only source for harm outcomes in relevant populations and over relevant time horizons.[67]

Treatment Information

Ideally, time-dependent information on treatment exposures should be considered, because patients may start and stop treatment during a study. Such information may only be available at the evidence generation stage, making it difficult for systematic reviews to include time-dependent information on treatment exposures.

Provision of a Benefit and Harm Comparison Metric That Considers Multiple Relevant Benefit and Harm Outcomes

Ideally, a quantitative benefit and harm assessment considers all relevant outcomes for benefit and harm and allows for the use of a benefit and harm comparison metric. Potential

metrics are QALYs, NNT/NNH ratio, or an event rate (e.g., rate of events prevented or caused by some treatment).[14]

Time of Occurrence of Benefit and Harm Outcomes

A quantitative benefit harm assessment should consider that the time of occurrence of benefit and harm outcomes may be different. Sometimes (e.g., for preventive treatments) harm outcomes are likely to precede benefit outcomes, whereas for other treatments, benefit might be experienced earlier than harm. In addition, a quantitative benefit harm assessment should ideally be able to incorporate variability of benefits and harms in relation to time.

Handling Different Data Types

Quantitative benefit harm assessments should be able to handle different types of data (binary, recurrent, continuous, and time to event). Many quantitative approaches identified here focus on binary outcomes with or without consideration of time to event. They cannot express some health, or patient-important, outcomes (e.g., quality of life or symptoms) appropriately as binary outcomes without substantial loss of information.

Uncertainty Estimates for the Benefit and Harm Comparison Metric

Uncertainty regarding the benefit and harm comparison metric is likely to be of key importance for decisionmakers and organizations making treatment recommendations. Typically, uncertainty is represented by 95% confidence intervals, but graphical displays such as the risk-benefit contour may express the extent of uncertainty.

As explained earlier in this report, decisionmakers will likely correlate benefit and harm outcomes, and they should take into consideration the potential interdependence of benefits and harms. Uncertainty may also arise from low-quality evidence. Other ways exist to express the uncertainty about the quality of evidence, such as the evidence grading schemes of the U.S. Preventive Services Task Force or the GRADE working group.

Incorporation of Preferences

A quantitative benefit harm assessment should explicitly state if and how preferences are incorporated. Most of the quantitative approaches identified here do not incorporate preferences, at least not explicitly. In the context of a systematic review, one can gather explicit data on patient preferences through identification of studies that used conjoint analysis or other methods to elicit patient preferences.

Consideration of Different Patient Profiles

Quantitative approaches like the NNT/NNH ratio, Gail/NCI or Net Clinical Benefit explicitly consider outcome risks. It is likely that the benefit and harm comparison for a drug like aspirin varies substantially according to age and gender since age and gender are associated with the risk for certain outcomes (e.g., myocardial infarction or gastrointestinal bleeding). Therefore, a quantitative benefit harm assessment should take into consideration different patient profiles.

Communication of Benefit Harm Assessment to Decisionmakers

Studies need to communicate a quantitative benefit harm assessment effectively to decisionmakers so that they can make informed decisions that are in line with their preferences. It is unknown which of the presentation formats are most effective. However, important aspects need to be considered, including: (1) parsing of information that illustrates a transparent and reproducible reduction of multidimensionality, (2) an explicit statement regarding whether preferences are already incorporated into the estimate or how individual preferences can be incorporated, and (3) graphical displays to convey quantitative information including uncertainty.[68] A recent paper highlights that, in contrast to most systematic reviews citing benefit and harm in different locations, a conceptual simple visualization of benefit and harm in a single image could enhance communication of the benefit and harm to decisionmakers.[68]

Influence of Values and Preferences in Assessing Benefits and Harms

General Considerations and Definitions

Decisions, whether at the policy or individual level, incorporate best evidence, patient information (e.g., disease severity, life-expectancy, or comorbidity), and patient preferences.[9,69] This statement, derived from the Effective Health Care Program's methods guide and the work of David Sackett, makes intuitive sense in the context of systematic reviews.

Decision analysis calls for all available evidence to be used in decisionmaking. Our team posits that preferences of patients and choices of those who generate evidence, synthesize it, make policy-level decisions, and translate evidence into practice, fundamentally inform the process of finding and using "best evidence." Preferences and choices affect every step of this process, from the generation of evidence to the application of evidence in clinical practice. Similarly, it is difficult to arrive at a benefit harm assessment based on "best evidence" without acknowledging preferences and choices. Many approaches have been used for quantitatively assessing benefit and harm in individual studies that may be relevant for evidence syntheses, such as systematic reviews and modeling tools. These approaches depend on the choices made by investigators and systematic reviewers about the incorporation of preferences.

Our team believes that we will arrive at better assessments of benefits and harms if we are explicit and transparent about the ways choices and preferences affect how we generate and synthesize evidence. For example, when trial investigators select outcomes, they implicitly make a choice that reflects what they think is important for decisionmaking. The same applies to systematic reviews, even if systematic reviewers may not always be explicit about the role of patient preferences. When systematic reviewers seek input from stakeholders (or key informants) about the key questions to address in a systematic review, the preferences of the key informants will affect this process. The key informants may have different preferences, as noted in the example of the reports on venous thromboembolism. [30,31] In these examples, pulmonary embolism was considered a clinically important outcome by both, but prevention of deep vein thrombosis was not considered clinically important by both.[30,31] To make users of systematic reviews (e.g., payers and guideline developers) aware of the role that choices and preferences play in systematic reviews, it is necessary to be explicit about the rationale for the decisions that were made in designing and conducting each review. In addition, to optimally assess and present data on benefits and harms, systematic reviewers should consider how end users interpret these systematic reviews.

Our objective in this section is to describe how preferences and choices play a role in various approaches to assessing the benefits and harms of interventions. The team reports its findings according to four stages of the process of translation: (1) evidence generation (e.g., RCTs and observational studies); (2) evidence synthesis (e.g., systematic reviews and meta-analyses); (3) processes used in moving from evidence generation to development of evidence-based medicine tools (e.g., modeling or simulation based on a decisionmaking context); and (4) the generation of evidence-based medicine tools (e.g., clinical practice guidelines and decision aids). We chose this organizational scheme because it is important to consider the full continuum of areas where

preferences and choices may influence systematic reviews and quantitative approaches to assessing the benefits and harms of interventions.

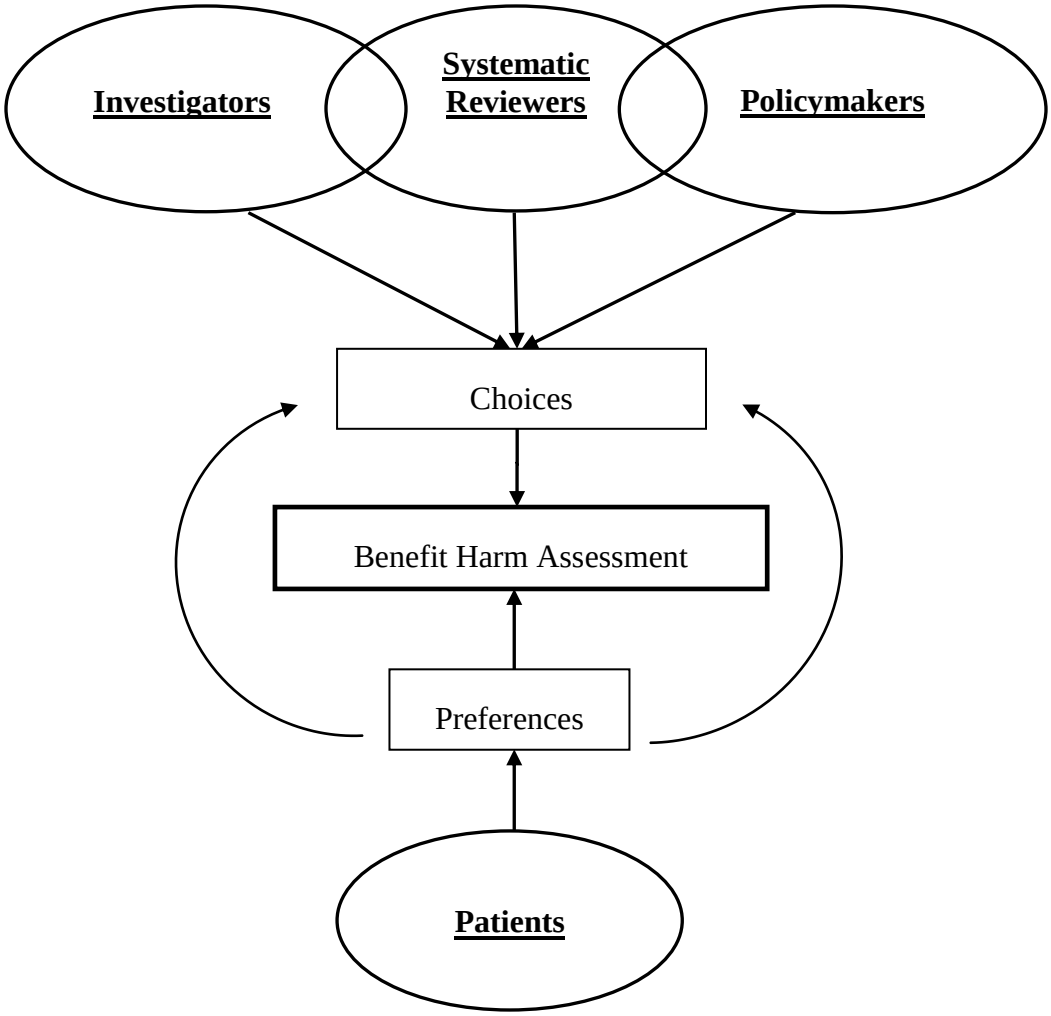
Definitions

Medical literature uses the terms "patient preferences" and "patient values" together or interchangeably in the area of benefit harm assessment. Often, the literature does not adequately define these terms and the difference between values and preferences is not clear.[70]

The Journal of the American Medical Association (JAMA) user's guide to the medical literature defines patient preferences and values as: "an overarching term that includes patients' perspectives, beliefs, expectations, and goals for health and life. We (JAMA) also use this phrase, more precisely, to mean the processes that individuals use in considering the potential benefits, harms, costs, and inconveniences of the management options in relation to one another."[70]

As has been done by others,[7] we believe distinguishing between values and preferences is useful. For purposes of this discussion, we have distinguished between patient values and patient preferences, as defined below, and we have added consideration of the choices that are made by decisionmakers when assessing benefits and harms of medical interventions (see Figure 4, Preferences and choices influencing decisionmaking at the policy level). This report refers predominantly to patient preferences and decisionmaker choices according to the working definitions given in the text below.

Figure 4. Preferences and choices influencing decisionmaking at the policy level[a]



aAs this report is focused on decisionmaking at the population level, we have not included providers of clinical care in this figure, although clinical providers are likely to have important roles in research, evidence synthesis, and policymaking. This figure does not represent individual level decisionmaking.

Patient Preferences

Patient preferences express the relative importance that patients or potential patients place on various health outcomes. They refer to the degrees of subjective satisfaction, distress, or desirability that patients or potential patients associate with a particular health outcomes. The various health states can be different severities of one condition (e.g., mild vs. moderate vs. severe dyspnea) or different conditions (e.g., hospital admission vs. severe dyspnea). Preferences are a consequence of values and beliefs and the specific contexts in which patients face decisionmaking. Utility instruments can both elicit and express preferences.

Patient Values

Values are a person's beliefs, desires, and expectations of what is right or wrong. Studies generally base values on predispositions as well as on family and cultural context and experiences. Values are latent traits that are not directly observable (i.e., measurable). Studies can only approximate values by how people express them (e.g., through preferences). Values are not specific to a certain context.

Decisionmaker Choices

Multiple decisionmakers may be involved in benefit and harm assessment, as described in Figure 4, including: investigators for clinical trials and observational studies (hereafter referred to as investigators*)*; systematic reviewers; policymakers such as payers and guideline developers; and ultimately, both clinicians and patients. Importantly, for stages I to III of benefit harm assessment, (described earlier in this report in Figure 1) clinician (provider) preferences may also factor in through their roles on guideline panels and regulatory bodies. End users of systematic reviews are typically policymakers and other decisionmakers, but may also include clinicians.

Ideally, all of these decisionmakers have patients as their primary focus and therefore care about patient preferences, but decisionmakers may need to make decisions in the absence of perfect information.[71] This is one of the reasons why minimizing conflicts of interest, or at least openly declaring them, is so important for benefit harm assessment.[72] Since patient preferences are often not available, these decisionmakers must operationalize what is known (or assumed to be true when evidence about patient preferences and their variability is lacking) about patient preferences, and they may incorporate other types of information into their choices as well. These include perceptions about societal preferences as well as recognition of varied preferences among specific subgroups.

Increasingly, the health care community is encouraging policymakers, such as guideline developers, to be explicit about the role that assumptions about patient preferences play in the process of issuing recommendations in practice guidelines.[26,73,74] Decisionmaker choices are important, and should be both explicit and transparent. For example, financial considerations might influence choices that trial investigators make about specific outcomes to measure. Similarly, investigators make choices regarding length of the study, outcomes followed, and eligibility criteria of a trial. For clarity, this report uses the term "investigator choices" to describe choices made by investigators that are informed by both patient preferences and other considerations. Similarly, we use the terms "systematic reviewer choices" for choices made by reviewers that are informed by the preferences of patients and other considerations, and "policymaker choices" for choices made by policymakers that are informed by the preferences of patients and other considerations.

Role of Choices and Preferences in Benefit Harm Assessment

The Role of Choices and Preferences in Evidence Generation

Investigator choices influence evidence generation in randomized controlled trials and observational studies in several ways. The choices are often implicit, and may or may not directly account for patient preferences.

Using the Population, Intervention, Comparator, Outcome, Timeframe, and Setting (PICOTS) framework, our team first looked at the many ways investigator choices affect study design. For example, investigator choices affect decisions regarding the definition of a study population, in terms of eligibility criteria and the aggressiveness with which the study recruits older, complex, or otherwise vulnerable populations. Investigator choices also influence decisions regarding the intervention to be tested, the length of the study, and the comparator(s).

Investigator choices also affect decisions regarding the selection of outcomes to be assessed in a clinical trial or observational study. For example, what is the primary outcome, and what are secondary outcomes? To consider the patient's perspective, investigators should choose outcomes important to patients (e.g., mortality or function) as primary outcomes, but in many cases researchers choose an intermediate, or surrogate, outcome for reasons of feasibility (e.g., short-term trials, smaller sample sizes). How closely linked surrogate outcomes are to health, or patient-important, outcomes varies by condition as well as by specific types of treatment, and often evidence is insufficient to appraise this linkage.[23,71,75] For example, to validate surrogate outcomes of oral agents for type 2 diabetes (e.g., glycated hemoglobin), systematic reviews should appraise the linkage between glycated hemoglobin and cardiovascular outcome separately for each drug class (such as metformin and thiazolidinediones).[23] The evidence that metformin lowers glycated hemoglobin and potentially reduces cardiovascular risk does not validate the use of thiazolidinediones to lower cardiovascular risk. Even though thiazolidinediones also lower glycated hemoglobin, similar to metformin, they may have a neutral or even adverse effect on cardiovascular outcomes.[76] Therefore, critical choices need to be made by investigators. Investigator choices also affect how study procedures account for outcomes that may be unintended consequences (such as harms).

Our team also looked at how investigator choices influence decisions regarding the analysis of a study. For example, investigator choices influence decisions regarding whether and how to assess for heterogeneity of treatment effects across study subpopulations. Investigator choices may also determine: (1) whether investigators conduct a quantitative benefit harm assessment and which outcomes enter such an assessment, (2) whether investigators use intention–to-treat or per-protocol analysis, (3) how rigorously investigators will ascertain data on harms, and (4) how frequently they will collect data on harm from participants who withdraw from the study. Participants may withdraw because of harms, and not counting these participants may result in the under-ascertainment of harms.[77]

The Role of Choices and Preferences for Evidence Synthesis

When synthesizing evidence, systematic reviewers attempt to incorporate different perspectives by soliciting input from diverse stakeholders or technical experts (e.g., generalist physicians, specialists, and patient representatives). These diverse inputs are implicitly assumed to provide information about the variability of preferences and values, but this is often an unverifiable assumption.

"Patient preference" was only introduced as a Medical Subject Headings term in PubMed in the year 2010, and the underlying information base for patient preferences is still evolving.[78] Thus, few evidence reports mention an explicit search for information on patient preferences. For example, in contrast to the large number of studies that reported on the effectiveness of medications for type 2 diabetes, very few studies are available on the role of patient preferences in weighing various outcomes in type 2 diabetes.

In evidence syntheses, systematic reviewers and the stakeholders they choose to engage make choices on the basis of implicit assumptions about preferences. The choices affect each component of the PICOTS framework for defining key questions. The decisions are parallel to those described in this report under "evidence generation." In the EPC Program, systematic reviewers make these decisions as they develop their key questions with the input of key informants including prominent investigators. Thus, preferences and choices of investigators and other stakeholders may influence what outcomes systematic reviewers consider important in the systematic review. The preferences and choices of stakeholders may also influence the assumptions that systematic reviewers make about the relative weights of benefits and harms.

An illustrative example is the distinction between reporting number needed to treat (NNT) and number needed to harm (NNH), and reporting a NNT/NNH ratio. A NNT/NNH ratio, unless explicitly calculated using relative weights, may implicitly assume that a given benefit and a given harm are equally important, although a decisionmaker could decide to compare the ratio to something other than one. Reporting NNT and NNH separately (e.g., no ratio) does not make any assumptions about the relative importance of the outcomes. This issue is related to the way in which the reviewer puts the results in context and helps the reader interpret them. There are extensions of the NNT/NNH ratio approach that explicitly consider patient preferences, or that consider whether the approach can handle more than one outcome for benefit and harm. For additional information, refer to the section in this report on quantitative approaches.

Quantitative approaches informed by preferences may bring worthwhile information to end users of systematic reviews regarding the balance of benefits and harms. Systematic reviewers should present the results of quantitative approaches for benefit harm assessment under different assumptions about patient preferences to help decisionmakers understand the implications of varied preferences (Figure 4).

The Role of Choices and Preferences for Processes, From Evidence Generation to Development of Evidence-based Medicine Tools: Modeling or Simulation

Modeling and simulation are important tools for assessing benefits and harms; several quantitative approaches to benefit harm assessment include these tools. These approaches all require input about preferences, in the choices of important outcomes, time horizons, and patient profiles. Modeling or simulation for a specific decisionmaking context also requires an understanding of the relative weights of benefits and harms (ideally based on patient preferences) and assumptions about whether the distributions of benefits and harms are independent.

The Role of Choices and Preferences for Development of Evidence-Based Medicine Tools

The role of preferences in the development of guidelines and decision aids is increasingly explicit.[71,79,80] The formulation of recommendations in guidelines depends on the ability of the guideline developers to reach conclusions about the uncertainty or variability of preferences on the behalf of patients. For example, if reviewers conclude that patient preferences vary only slightly and that the balance of benefit and harm favors benefits, guideline developers are likely to issue a strong recommendation if it is also supported with high-quality evidence. Uncertainty or variability in values and preferences may lead to a weak recommendation.[26]

Understanding when patient preferences vary is therefore important for determining the strength of the recommendation in guidelines. Decisionmakers often have imperfect information about patient preferences, but nevertheless must arrive at a decision that incorporates perceived preferences. Thus, decisionmakers must make inferences about patient preferences that are, in turn, incorporated into the decisionmaking process. This role contributes to an emerging consensus about the necessity to minimize conflicts of interest and to ensure that the leadership of guideline panels includes those without conflicts of interest. A conflict of interest in this case would be a situation in which a decisionmaker could benefit from a recommendation that is not reflective of patient preferences.

Similar issues affect decision aids.[81] To create useful decision aids, decisionmakers must identify (early in the development process) the appropriate clinical situations where preferences affect decisionmaking. For example, decisionmakers need to recognize how patients view the effects of urinary incontinence on their quality of life and incorporate preferences regarding this outcome into a decision aid about treatment options.[82]

Use of Evidence-Based Medicine Tools in Clinical Practice

Clinical decisionmakers must combine individual patient preferences with guideline recommendations and decision aids. Population-level decisions, such as those made to inform coverage or guideline recommendations, will not be identical to every individual's decision, because both the context of individual decisions and the preferences of individuals may be different.[83] This is true even in settings with individual patients having shared decisionmaking using high-quality, applicable evidence regarding the likelihood of benefit and harm. This difference is due primarily to the variation in the relative weights patients would apply to the different possible outcomes, including benefits and harms. Modern decision aids often include an instrument to elicit preferences and to make suggestions to patients that are based on their own preferences.[84]

Patient preferences may affect decisionmaking to varying degrees. Most health care decisions are preference-sensitive, especially in the context of chronic diseases.[85,86] Preference-sensitive decisions are made in situations where different decisionmakers may reach different conclusions based on their personal preferences for what is known about the benefit and harm. For example, Protheroe et al. showed that patient preferences have a strong effect on whether or not guideline-concordant care is selected for the prevention of stroke from atrial fibrillation with aspirin or warfarin.[83] Similarly, Sussman et al. demonstrated that patient preferences have a large impact on whether or not aspirin would be recommended for the primary prevention of myocardial infarction or stroke.[87] Exceptions occur when preferences do not play much of a role because of unambiguous evidence that a certain health intervention is virtually always necessary. Examples would be surgical repair of ongoing bleeding, displaced fracture due to trauma, or antibiotic treatment of a child with sepsis.

Principles for Assessing Benefit and Harm in Systematic Reviews

Systematic Review Protocol Development

Identify the Key Potential Benefits and Harms

Systematic reviews should attempt to inform the harm and benefit assessment by identifying the key benefits and harms, including the health outcomes most important to patients, in the analytic framework. The Agency for Healthcare Research and Quality (AHRQ) Methods Guide for Comparative Effectiveness Reviews lists this as standard guidance for systematic reviews.[9] Systematic reviewers should carefully consider how they will combine the effects of treatments on multiple benefits and harms as well as how they will incorporate patient preferences for these different outcomes.

When data on important health outcomes are unavailable, reviewers should report on the way in which treatments affect potential surrogate outcomes and assess the strength of the linkage between surrogate measures and either benefits or harms. In circumstances where only benefits on potential surrogate outcomes are available and their linkages to health outcomes are not validated, reviewers should discuss with end users the potential value of a benefit harm assessment on surrogate outcomes.

Report the Characteristics and Assumptions of the Selected Quantitative Approaches

The various quantitative approaches differ in key characteristics. While most quantitative approaches for benefit harm assessment are feasible whether using aggregate or individual patient data, some methods (e.g., benefit-less-risk analysis) specifically require individual patient data. Some are typically based on RCTs only (e.g., number-needed-to-treat to number-needed- to harm ratio (ratio of NNT over NNH), while others can consider both experimental and observational data (e.g., probabilistic simulation modeling, Gail/NCI (National Cancer Institute) approach, or multi criteria decision analysis).

Another key characteristic of a quantitative approach for benefit harm assessment is the number and diversity of benefit and harm outcomes that it incorporates. While some approaches typically focus on a single or a few outcomes of similar severity (e.g., NNT/NNH ratio or transparent uniform risk benefit overview), others incorporate a potentially large number of outcomes of different importance (e.g., Gail/NCI approach or multi criteria decision analysis). Since little empirical data exist to suggest that one particular quantitative approach for assessing benefits and harms is superior to another, systematic reviewers should consider which quantitative approach is most appropriate for the decisionmaking context, the data available through the systematic review, and the methodological expertise of the review team. If the reviewers decide for or against a particular approach to quantitative assessment of benefit and harm balance, they should report the rationale for this decision. They should also describe the characteristics and assumptions of the selected quantitative approach.

State Whether Preferences Were Considered in the Benefit Harm Assessment, and If So, Describe How These Were Ascertained and How Variation in Preferences Would Affect the Assessment

Preferences affect the assessment of benefits and harms in any review of evidence. In current practice, these preferences are often implicit and those conducting an assessment of benefits and harms do not necessarily recognize and transparently report them. This is true whether a review simply reports on benefits and harms, or whether quantitative approaches are used to provide a benefit and harm comparison estimate.

Systematic reviews should explain and justify the rationale for their choice of various study designs or various outcomes that inform the assessment of benefits and harms and any other steps in the process of evidence synthesis where preferences play a role. They should consider performing systematic searches for studies on patient preferences for relevant outcomes and conducting sensitivity analyses to determine the impact of varying preferences on assessing benefit and harm.

Describe Whether Systematic Reviewers Use a Qualitative Assessment or a Quantitative Approach for Benefit Harm Assessment

Reviewers may choose to conduct a qualitative assessment of the results of a systematic review for benefit and harm outcomes. These results may or may not be summarized using meta-analysis. A qualitative assessment is currently the most common approach for assessing the balance of benefits and harms by systematic reviewers and policymakers. A quantitative assessment using a benefit and harm comparison metric may be informative for the third stage of decisionmaking (see Figure 1).

If systematic reviewers use a quantitative approach, they should say so explicitly. Systematic reviews should clearly explain whether or not they choose to use a quantitative approach that will provide a benefit and harm comparison metric. If a quantitative approach is chosen, the systematic review protocol should outline which approach is chosen and the justification behind the choice. The choice of a particular approach has implications for: (1) the type of evidence (e.g., randomized trials and/or observational studies) that needs to be identified; (2) the electronic and/or non-electronic search strategies; (3) the involvement of stakeholders and if their preferences need to be considered; (4) the type of outcome data to be extracted or requested from primary studies; and (5) the methodological-statistical expertise needed to conduct the analyses.

Conduct and Reporting of Systematic Reviews

Preserve Information When Reporting on Benefit and Harm

An information-preserving approach allows the users of systematic reviews to calculate any metric that compares the event rates in two treatment groups. For example, if reviewers present the event rate (e.g., number of events per 1,000 person-years) of an outcome for each treatment group, it allows users to calculate both risk difference and relative risk and is an information-preserving approach. If reviewers only presented the relative risk, it would not be possible to obtain the risk difference. This is reiterated in the AHRQ Methods Guide for Comparative Effectiveness Reviews, which advises that reviewers should report absolute risks along with

relative risks.[9] Systematic reviewers should conduct the quantitative assessment of benefits and harms on an absolute scale or use both absolute and relative scales. Another example of this principle is in how reviewers combine outcomes. An information-preserving approach would report the event rates for each individual outcome in each treatment group. Combining all the benefit outcomes into a single outcome (e.g., a composite outcome) and only reporting this composite outcome is not information-preserving.

State How Decisions About Comparisons, Outcomes, Baseline Risks, and Time Horizons Were Made To Increase Transparency.

To enable end users to replicate the methods, systematic reviewers should state how they made decisions about comparisons, outcomes, baseline risks, and time horizons to increase transparency of reporting on benefits and harms. Systematic reviews should aim to consistently report the time horizon of included studies. Reviewers also need to provide the sources of data and assumptions used in conducting a benefit harm assessment.

For example, if reviewers use the NNT and NNH approach, systematic reviews should not only provide information on the underlying meta-analysis for the intervention that informed the relative or absolute risk estimates, but also information on whether or not the baseline rate used was the control event rate in the trials or some other population-based study. The reviewers need to indicate the time periods that relate to the estimates. Any additional assumptions about estimating NNT and NNH should be clarified. Such assumptions could include the assumption that benefits and risks are constant over time. Some of these assumptions are unverifiable and inherent to the respective quantitative approach. For example, some methods require the assumption that effect estimates do not vary across populations. On the other hand, assumptions about constant risks over time, or variations in the contour of benefit and harm over time, could be verified with either individual patient data or closer examination of summary data. Assumptions should be justified to enable end users to determine whether the assumptions fit their needs and whether the source of the estimates is applicable to their context.

Convey Sampling Uncertainty and Uncertainty in the Strength of the Evidence.

Systematic reviewers should qualitatively describe the various sources of uncertainty surrounding the balance of a benefit harm assessment. Reviewers should be as explicit as possible concerning the applicability of the evidence on harm to particular subgroups such as older adults or people with important comorbidity.

Previous work has noted that uncertainty can arise from at least five different types of issues: inherent uncertainty regarding the future, uncertainty regarding validity, uncertainty regarding significance for the individual, uncertainty related to complexity (e.g., that the risks for several outcomes each change with time), or uncertainty related to what is not known.[88] Reviewers should pay particular attention to sampling uncertainty and report not only point estimates of effect, but also report confidence intervals around benefit and harm estimates as recommended in the AHRQ Methods Guide for Comparative Effectiveness Reviews. To convey uncertainty around the strength of evidence, reviewers should provide an explicit strength of evidence grade for each important benefit and each important harm. Probabilistic analysis and sensitivity analysis are two examples of quantitative approaches that reviewers can use to explore uncertainty regarding the validity of effect estimates. Reviewers may not have enough

information to estimate the potential effects of all sources of uncertainty, but they should at least acknowledge each source of uncertainty. Decisionmakers then can make decisions with better awareness of what is not known with certainty, such as the long-term effects of treatments.

Discussion

In the pages that follow, we discuss the key findings for each of our objectives. We then discuss the limitations of this report, identify future research needs, and summarize our overall conclusions.

For Objective 1, we found a number of challenges related to the conduct of a systematic review, the grading of the strength of evidence, and the incorporation of preferences in a review.

First, benefit and harm, or uncertainty about the balance between the two, may vary for subgroups of the population. Assessment of benefit and harm may depend on the linkages between surrogates and health outcomes for specific interventions, and whether these linkages vary across subgroups. The time horizon for studying benefits and harms may be inadequate in many studies included in a review. In addition, the fidelity with which the intervention was applied may vary across studies, making a synthesis of benefits and harms difficult.

Another significant challenge is to judge the applicability of studies to the target population of interest with respect to both benefits and harms. Studies may be designed to provide more robust data on benefits than on harms, requiring the assessment of harms through study designs other than randomized controlled trials (RCTs), such as observational studies or case reports. When such information asymmetry exists, the strength of evidence will vary for different benefits and harms, making it difficult to rate the strength of evidence for the overall balance of benefits and harms.

Using a quantitative approach for assessment of benefits and harms may require several assumptions. Data are usually unavailable on the joint distribution of benefits and harms under various scenarios, and it may vary among different patient profiles. Benefits and harms may be reported on different scales. Values and preferences affect how people weigh the relative importance of benefit and harm outcomes. Such data on values and preferences are usually unavailable to systematic reviewers.

In Objective 2, we described the methodological characteristics of existing quantitative approaches for assessing benefit and harm. The main findings of our review of the available quantitative approaches for assessing benefits and harms using a unifying framework are summarized below.

(1) Existing quantitative approaches can be categorized into approaches that consider single or multiple benefit and harm outcomes as well as those approaches that use or do not use a benefit and harm comparison metric.

(2) Although none of the approaches seemed to be developed specifically for handling aggregate data in systematic reviews, systematic reviews can use all quantitative approaches except for benefit-less-risk analysis, Boers' method, and stated preference method or maximum acceptable risk, all of which require individual patient data that may include data from randomized trials (benefit-less-risk analysis) but also data from preference-eliciting surveys.

(3) None of the quantitative approaches for assessing benefit and harm explicitly consider the asymmetry in the quality and quantity of evidence, which is generally higher and larger, respectively, for benefits as compared with harms. Such asymmetry is particularly important in the context of a systematic review because it has implications for the search strategy, selection of the evidence, and the overall workload of reviewers.

(4) No single quantitative approach can be favored clearly over the others in the context of every systematic review. The selection of a quantitative approach depends on the number of key outcomes, the need or desire for a benefit and harm comparison metric, the balance between

offering as many of the outlined desired properties as possible and the feasibility of ascertaining a comprehensive evidence base, and the availability of epidemiological and statistical expertise. Finally, a comprehensive benefit harm assessment could use a combination of the approaches reviewed here.

The team identified a number of assumptions that researchers make implicitly when applying almost any quantitative approach for assessing benefits and harms. First, for most approaches, researchers assume that one or more benefit and harm outcomes can be put on the same scale to calculate a benefit and harm comparison metric. Challenges for putting different outcomes on the same scale include their relative importance to decisionmakers (which may require different weighting), simplification of the outcomes (e.g., dichotomizing continuous outcomes, which may lead to substantial loss of information), or different methods and timing used in ascertaining different outcomes. A benefit and harm comparison metric may offer some advantages particularly in the context of situations where multiple outcomes are important and where patient, provider, and policymaker preferences vary. The advantage of using such a metric, compared to an approach without a benefit and harm comparison metric, is that researchers make explicit assumptions about the relative importance of outcomes. Sensitivity analyses can provide evidence showing how the benefit and harm balance changes if researchers make different assumptions. A single estimate may provide some advantages for the communication of net treatment benefits to patients and avoids overwhelming the patients with data on multiple different outcomes. Ultimately, decisionmakers should judge the usefulness of a benefit and harm comparison metric for the types of decisions they need to make.

Second, the use of multiple endpoints generally seems more feasible than the use of composite outcomes in quantitative approaches for assessing benefits and harms. Reviewers may assume they can build composite outcomes that provide useful information for decisionmakers. Some reviewers may see this as similar to putting different outcomes on the same scale. However, composite outcomes may not be a suitable option in systematic reviews, because benefits and harms are often defined differently across trials. Consistent composite outcomes can only be constructed if detailed data on individual endpoints are available from the primary studies. In certain instances, composite outcomes may provide misleading results.[22,89]

Third, the reports using these quantitative approaches did not consider the joint distribution of benefit and harm even when individual patient data were available. Uncertainty estimates for the benefit and harm comparison metric may be different if studies considered joint distributions. We did not identify studies that quantified such a difference between uncertainty estimates, with and without consideration of the joint distributions. Systematic reviews usually do not have access to information on the joint distribution of benefit and harm outcomes in the primary studies. Thus, systematic reviewers are unable to consistently report the joint distribution of benefit and harm outcomes across studies. However, systematic reviewers should consider that ignoring the joint distribution when interpreting results from a quantitative benefit harm assessment could present a significant limitation.

If systematic reviewers decide to conduct a quantitative benefit harm assessment, they may choose among the approaches presented in this report, or combinations thereof and among others outside our scope, like decision analysis. Furthermore, it is likely that additional quantitative approaches are currently being developed or that existing approaches are being modified. Feasibility is obviously one of the factors that will drive the decision on which approach to choose. Relatively simple quantitative approaches include number needed to treat (NNT) and number needed to harm (NNH), NNT/NNH ratio and its modifications, minimum clinical

efficacy, and transparent uniform risk-benefit overview. Other more complex quantitative approaches may be difficult to apply in most systematic reviews, such as incremental net health benefit, net clinical benefit, probabilistic simulation method, multicriteria decision analysis, or Gail/NCI (National Cancer Institute) because they require information that systematic reviews don't collect automatically, and epidemiological and statistical expertise that may not be available.

It is important to clearly define the decisionmaking context of a systematic review. This includes a characterization of the decisionmakers that need to be informed and their need to solve decisional conflicts. Defining the decisionmaking context helps define the exact research question, the eligibility criteria for studies and aspects of study design, and the quantitative approach for assessing benefit and harm that will fulfill as many of the desired properties as possible. As a result, it may be that for some decisionmaking contexts, available RCTs provide all the evidence needed and one of the simpler approaches for quantitative benefit harm assessment fulfills important desired properties. If the number of benefit and harm outcomes becomes larger, the time horizon longer, and the patient population more heterogeneous, it is likely that only one of the more complex approaches will provide a quantitative benefit and harm assessment that fulfills the described properties and meets the decisionmakers' needs.

Since none of the approaches are perfect and all require some assumptions, sensitivity analyses are important. When using a specific approach it may be wise to specify the assumptions made for the main analysis and to assess how the results change if the assumptions are modified.

Also, it may be valuable to consider additional approaches to assess whether the results and conclusions from a benefit harm assessment depend on the strengths and limitations of the specific quantitative approach used. An additional important consideration for systematic reviewers is that some of the approaches reviewed here may be combined. For example, NNT and NNH, as well as some of the visual approaches such as risk-benefit contour, could be used to present the results in a multicriteria decision analysis. Alternatively, simulation could be used if the study did not assess the benefit and harm outcomes in the full range of the population that clinicians might consider treating.

For Objective 3 we discussed the influence of patient values and preferences, as well as decisionmaker choices, in assessing benefits and harms across each step of a systematic review and the process of translation before and after systematic reviews. Choices and preferences affect benefit harm assessments across the entire path of evidence translation: evidence generation (e.g., RCTs and observational studies); evidence synthesis (e.g., systematic reviews and meta-analyses); processes used in moving from evidence generation to development of evidence-based medicine tools, such as modeling or simulation based on a decisionmaking context; and the generation of evidence-based medicine tools (e.g., clinical practice guidelines, decision aids). In current practice, choices are often implicit and not transparently reported in many of these steps. For systematic reviewers, this is true whether reviewers simply report on the benefits and harms, or use quantitative approaches to assess the benefits and harms. Thus, deciding not to use quantitative approaches does not mean that choices are not relevant. Choices and preferences also affect how guideline developers frame recommendations, how regulatory bodies make decisions at the population level, and how clinicians, patients, and other end users make decisions at the individual level.

For Objective 4, we formulated principles for assessing benefits and harms in systematic reviews that relate to protocol development or conduct and reporting. Because no method is

clearly superior to others for conducting an assessment of benefits and harms in a systematic review, the principles generally emphasize the need to report not just results but also to describe how decisions were made. We emphasized the need to report which assumptions related to chosen methods are likely to be critical. Since benefit harm assessments are highly multidimensional, there will always be some arbitrary decisions involved in the assessments. Sensitivity analyses are therefore important to quantify the effect of arbitrary decisions on the results of quantitative benefit harm assessments.

Estimating uncertainty is also important for decisionmaking. Also, it is critical to preserve information while reporting on benefits and harms to assess variation in benefit harm assessment across relevant subgroups. As is recommended in the Agency for Healthcare Research and Quality Methods Guide for Comparative Effectiveness Reviews, to overcome information asymmetry the assessment of harms in systematic reviews should not be restricted to specific study types.[9] While observational studies provide important information on harms associated with certain treatments, they are more susceptible to confounding, selection, information and reporting bias than RCTs.

Limitations

We understand that this report has several limitations. We reviewed a small sample of evidence reports, so it is possible that we did not identify all the relevant challenges. The sample of evidence reports we reviewed was illustrative of the major challenges facing systematic reviews in conducting a benefit harm assessment.

Since we did not conduct a systematic review, we may have missed some quantitative approaches to benefit and harm assessment. We reviewed the most commonly used approaches so it is possible that we did not identify all characteristics of benefit harm assessments that could be applied to a systematic review. We did not evaluate quantitative methods that researchers have used outside the health care setting. Decision trees and influence relevance diagrams may require data beyond what is typically available from evidence generated from a systematic review and were beyond the scope of this project. We did not evaluate the full range of other decision analytic approaches that use principles of risk analysis, mathematical psychology, evaluation sciences, and conjoint measurement theory. Our principles are not prescriptive and cannot address all the challenges of benefit harm assessment, which fall outside the domain of systematic reviews.

Future Research

Further empirical work is needed to improve understanding of the utility of different quantitative approaches for assessing benefits and harms in systematic reviews. Such work should assess the value of the quantitative approaches from the standpoint of decisionmakers. For example, what metrics of benefits and harms are most relevant and useful to various types of decisionmakers? How do decisionmakers feel about the desirability of combining multiple outcomes into one metric when the clinical intervention has a complex set of benefits and harms? How important is it to use metrics that incorporate preferences? How important is it to use metrics that are transparent about potential uncertainty? As methods continue to evolve for assessing and reporting on the balance of benefits and harms, it will be necessary to further develop the methods for conducting systematic reviews of preferences as well as for synthesizing evidence on preferences.[78] It will also be necessary to give more attention to the methods for

grading the strength of evidence when a quantitative approach is used to assess benefits and harms.

Future studies need to evaluate whether network meta-analysis can address some of the challenges identified above. Researchers have recently published network meta-analyses on the relative effects of alternative interventions on benefit and harm outcomes.[25] The lack of evidence from direct comparisons in the context of systematic reviews of benefit harm assessments can be evaluated in network meta-analysis. Such analyses provide estimates on the relative effectiveness of health care interventions by considering both direct and indirect evidence. Studies included in a network meta-analysis should be, as in any conventional meta-analysis, comparable in terms of quality (e.g., risk of bias), applicability, and measurement of outcomes. In addition, meta-analyses should include a close examination of the consistency of results from direct and indirect evidence.[90]

Conclusions

A number of quantitative approaches for benefit harm assessment are available for systematic reviewers. The choice of a particular approach depends on the decision to be informed, the available data, and the epidemiological-statistical expertise of the systematic review team. Quantitative approaches may often be attractive because they enhance transparency and help decisionmakers understand the multidimensionality of benefit harm assessments. By considering the challenges of benefit harm assessments, the characteristics of available quantitative approaches, and the recommended principles for using a quantitative approach, systematic reviewers should be able to choose an approach that will enhance their assessment of the balance of benefits and harms in each review.

References

1. Bennett WL, Wilson LM, Bolen S, et al. Oral diabetes medications for adults with type 2 diabetes: an update. Comparative Effectiveness Review No. 27. AHRQ Publication No. 11-EHC038-EF. Rockville, MD: Agency for Healthcare Research and Quality. March 2011. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm

2. Matchar DB, McCrory DC, Orlando LA, et al. Comparative effectiveness of angiotensin-converting enzyme inhibitors (ACEIs) and angiotensin ii receptor antagonists (ARBs) for treating essential hypertension [Internet]. Comparative Effectiveness Review No. 10. AHRQ Publication No. 08-EHC003-EF. Rockville, MD: Agency for Healthcare Research and Quality; November 2007.

3. Bravata DM , McDonald KM, Gienger AL, et al. Comparative Effectiveness of Percutaneous Coronary Interventions and Coronary Artery Bypass Grafting for Coronary Artery Disease. AHRQ Publication No. 08-EHC002-EF. Rockville, MD: Agency for Healthcare Research and Quality; October 2007.

4. Coleman CI, Baker WL, Kluger J, et al. Comparative effectiveness of angiotensin converting enzyme inhibitors or angiotensin II receptor blockers added to standard medical therapy for treating stable ischemic heart disease. Comparative Effectiveness Reviews No. 18. AHRQ Publication No. 10-EHC002-EF 2009. Rockville, MD: Agency for Healthcare Research and Quality; October 2009.

5. Bolen S, Wilson L, Vassy J, et al. Comparative Effectiveness and Safety of Oral Diabetes Medications for Adults With Type 2 Diabetes. Comparative Effectiveness Reviews No. 8. AHRQ Publication No. 07-EHC010-EF. Rockville, MD: Agency for Healthcare Research and Quality; July 2007. Available at: www.effectivehealthcare.ahrq.gov/reports/final.cfm

6. Hochman M, McCormick D. Characteristics of published comparative effectiveness studies of medications. JAMA. 2010; 303(10):951-8.

7. Froberg DG, Kane RL. Methodology for measuring health-state preferences--I: measurement strategies. J Clin Epidemiol :1989;42(4):345-54.

8. Institute of Medicine. Finding What Works in Health Care: Standards for Systematic Reviews. Eden J, Levit L, Berg A, et al., eds. Washington, DC: National Academies Press; 2011.

9. Methods Guide for Effectiveness and Comparative Effectiveness Reviews. AHRQ Publication No. 10(11)-EHC063-EF. Rockville, MD: Agency for Healthcare Research and Quality. March 2011. Chapters available at: www.effectivehealthcare.ahrq.gov.

10. Nissen SE, Wolski K. Effect of rosiglitazone on the risk of myocardial infarction and death from cardiovascular causes. N Engl J Med. 2007; 356(24):2457-71.

11. Bangalore S, Kumar S, Messerli FH. Angiotensin-converting enzyme inhibitor associated cough: deceptive information from the Physicians' Desk Reference. Am J Med. 2010; 123(11):1016-30.

12. Bennett WL, Maruthur NM, Singh S, et al. Comparative effectiveness and safety of medications for type 2 diabetes: an update including new drugs and 2-drug combinations. Ann Intern Med. 2011; 154(9):602-13.

13. Guo JJ, Pandey S, Doyle J, et al. A review of quantitative risk-benefit methodologies for assessing drug safety and efficacy-report of the ISPOR risk-benefit management working group. Value Health. 2010; 13(5):657-66.

14. Gail MH, Costantino JP, Bryant J, et al. Weighing the risks and benefits of tamoxifen treatment for preventing breast cancer. J Natl Cancer Inst. 1999; 91(21):1829-46.

15. European Medicines Agency. Report of the Committee for Medicinal Products Working Group on Benefit-Risk Assessment Models and Methods. EMEA/CHMP/15404/2007 2007. Available at: http://www.ema.europa.eu/docs/en_GB/documen_library/Regulatory_and_procedural_guideline/2010/01/WC500069668.pdf. Accessed October 21, 2012.

16. O'Neill RT. A perspective on characterizing benefits and risks derived from clinical trials: can we do more? Drug Information Journal. 2008; 42(3): 235-45.

17. Hughes DA, Bayoumi AM, Pirmohamed M. Current assessment of risk-benefit by regulators: is it time to introduce decision analyses? Clin Pharmacol Ther. 2007; 82(2):123-7.

18. Psaty BM, Furberg CD. COX-2 inhibitors--lessons in drug safety. N Engl J Med. 352. 2005:1133-5.

19. Caroll C, Patterson M, Wood S, et al. A conceptual framework for implementation fidelity. Implement Sci. 2007; 30(2):40.

20. Dane AV, Schneider BH. Program integrity in primary early and secondary prevention. Are implementation effects out of control? Clin Psychol. Rev. 1998 Jan;18(1):23-45.

21. Helfand M, Balshem H. AHRQ Series Paper 2: Principles for developing guidance: AHRQ and the effective health-care program. J Clin Epidemiol. 2010; 63. 2010:484-90.

22. Ferreira-Gonzalez I, Busse JW, Heels-Ansdell D, et al. Problems with use of composite end points in cardiovascular trials: systematic review of randomised controlled trials. BMJ. 2007; 334(7597):786.

23. Institute of Medicine Committee on Qualification of Biomarkers and Surrogate Endpoints in Chronic Disease. Evaluation of Biomarkers and Surrogate Endpoints in Chronic Disease. Micheel CM, Ball JR, eds. Washington, DC:The National Academies Press; 2010.

24. Hulley S, Grady D, Bush T. for the Heart and Estrogen/progestin Replacement Study (HERS) Research Group. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. JAMA. 1998; 280:605-13.

25. Chou R, Aronson N, Atkins D, et al. AHRQ Series Paper 4: Assessing Harms When Comparing Medical Interventions: AHRQ and the Effective Health-care Program. J Clin Epidemiol. 2010; 63(5):502-12.

26. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: An Emerging Consensus on Rating Quality of Evidence and Strength of Recommendations. BMJ. 2008; 336(7650):924-6.

27. Sin DD, Tashkin D, Zhang X, et al. Budesonide and the risk of pneumonia: a meta-analysis of individual patient data. Lancet 2009; 374(9691):712-9.

28. Greenland S. Additive Risk versus Additive Relative Risk Models. Epidemiology. 1993; 4:32-6.

29. Akl EA, Oxman AD, Herrin J, et al. Using alternative statistical formats for presenting risks and risk reductions. Cochrane Database of Systematic Reviews 2011(3): CD006776.

30. American Academy of Orthopedic Surgeons. Clinical Guideline on Prevention of Pulmonary Embolism in Patients Undergoing Total Hip or Knee Arthroplasty. [Web Page]. May 2007; Available at http://www.aaos.org/research/guidelines/pe_guideline.pdf .

31. Geerts WH, Bergqvist D, Pineo GF, et al. American College of Chest Physicians. Prevention of venous thromboembolism: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Chest. 2008; 133(6 Suppl):381S-453S.

32. Lacasse Y, Goldstein R, Lasserson TJ, et al. Pulmonary rehabilitation for chronic obstructive pulmonary disease. Cochrane Database Syst Rev. 2006 (4):CD003793.

33. Singer DE, Albers GW, Dalen JE, et al. Antithrombotic therapy in atrial fibrillation: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines (8th Edition). Chest. 2008; 133(6 Suppl):546S-92S.

34. Dolan JG. Multi-criteria Clinical Decision Support: A Primer on the Use of Multiple Criteria Decision Making Methods to Promote Evidence-based, Patient-centered Healthcare. Patient. 2010; 3(4):229-48.

35. Singh S, Dolan JG, Centor RM. Optimal management of adults with pharyngitis--a multi-criteria decision analysis. BMC Med Inform Decis Mak. 2006; 6:14.

36. Fisher B, Costantino JP, Wickerham DL, et al. Tamoxifen for the prevention of breast cancer: current status of the National Surgical Adjuvant Breast and Bowel Project P-1 study. J Natl Cancer Inst. 2005; 97(22):1652-62.

37. Shakespeare TP, Gebski VJ, Veness MJ, et al. Improving interpretation of clinical studies by use of confidence levels, clinical significance curves, and risk-benefit contours. Lancet. 2001; 357(9265):1349-53.

38. Chong J, Poole P, Leung B, et al. Phosphodiesterase 4 inhibitors for chronic obstructive pulmonary disease. Cochrane Database of Systematic Reviews. 2011 (5): CD002309. DOI: 10.1002/14651858.CD002309.pub3.

39. Aspirin for the prevention of cardiovascular disease: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med. 2009; 150(6):396-404.

40. Chuang-Stein C. A new proposal for benefit-less-risk analysis in clinical trials. Control Clin Trials. 1994; 15(1):30-43.

41. Boers M, Brooks P, Fries JF, et al. A first step to assess harm and benefit in clinical trials in one scale. J Clin Epidemiol. 2010; 63(6):627-32.

42. Ponce RA, Bartell SM, Wong EY, et al. Use of quality-adjusted life year weights with dose-response models for public health decisions: a case study of the risks and benefits of fish consumption. Risk Anal. 2000; 20(4):529-42.

43. Garrison LP Jr, Towse A, Bresnahan BW. Assessing a structured, quantitative health outcomes approach to drug risk-benefit analysis. Health Aff (Millwood). 2007 May-Jun;26(3):684-95.

44. Mussen F, Salek S, Walker S. A quantitative approach to benefit-risk assessment of medicines - part 1: the development of a new model using multi-criteria decision analysis. Pharmacoepidemiol Drug Saf. 2007; 16 Suppl 1:S2-S15.

45. Mussen F, Salek S, Walker S, et al. A quantitative approach to benefit-risk assessment of medicines - part 2: the practical application of a new model. Pharmacoepidemiol Drug Saf. 2007; 16 Suppl 1:S16-41.

46. Holden WL, Juhaeri J, Dai W. Benefit-risk analysis: a proposal using quantitative methods. Pharmacoepidemiol Drug Saf. 2003; 12(7):611-6.

47. Dolan JG. Shared decision-making--transferring research into practice: the analytic hierarchy process (AHP). Patient Educ Couns. 2008; 73(3):418-25.

48. Holden WL, Juhaeri J, Dai W. Benefit-risk analysis: examples using quantitative methods. Pharmacoepidemiol Drug Saf. 2003; 12(8):693-7.

49. Sutton AJ, Cooper N, Abrams KR, et al. A Bayesian approach to evaluating net clinical benefit allowed for parameter uncertainty. J Clin Epidemiol. 2005; 58(1):26-40.

50. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med. 1988; 318(26):1728-33.

51. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. BMJ 1995; 310(6977):452-4.

52. Osiri M, Suarez-Almazor ME, Wells GA, et al. Number needed to treat (NNT): implication in rheumatology clinical practice. Ann Rheum Dis. 2003; 62(4):316-21.

53. Walter SD, Sinclair JC. Uncertainty in the minimum event risk to justify treatment was evaluated. J Clin Epidemiol. 2009; 62(8):816-24.

54. Lynd LD, Najafzadeh M, Colley L, et al. Using the incremental net benefit framework for quantitative benefit-risk analysis in regulatory decision-making—a case study of alosetron in irritable bowel syndrome. Value Health. 2010; 13(4):411-7.

55. Shaffer ML, Watterberg KL. Joint distribution approaches to simultaneously quantifying benefit and risk. BMC Med Res Methodol. 2006; 6:48.

56. Lynd LD, O'brien BJ. Advances in risk-benefit evaluation using probabilistic simulation methods: an application to the prophylaxis of deep vein thrombosis. J Clin Epidemiol. 2004; 57(8):795-803.

57. Gelber RD, Goldhirsch A, Cole BF. Evaluation of effectiveness: Q-TWiST. The International Breast Cancer Study Group. Cancer Treat Rev. 1993; 19 Suppl A:73-84.

58. Gelber RD, Goldhirsch A, Cole BF, et al. A quality-adjusted time without symptoms or toxicity (Q-TWiST) analysis of adjuvant radiation therapy and chemotherapy for resectable rectal cancer. J Natl Cancer Inst. 1996; 88(15):1039-45.

59. Sherrill B, Amonkar MM, Stein S, et al. Q-TWiST analysis of lapatinib combined with capecitabine for the treatment of metastatic breast cancer. Br J Cancer. 2008; 99(5):711-5.

60. Ryan M, McIntosh E, Shackley P. Methodological issues in the application of conjoint analysis in health care. Health Econ. 1998; 7(4):373-8.

61. Johnson FR, Ozdemir S, Mansfield C, et al. Are adult patients more tolerant of treatment risks than parents of juvenile patients? Risk Anal. 2009; 29(1):121-36.

62. Hauber AB, Mohamed AF, Johnson FR, et al. Treatment preferences and medication adherence of people with Type 2 diabetes using oral glucose-lowering agents. Diabet Med. 2009; 26(4):416-24.

63. Johnson FR, Van Houtven G, Ozdemir S, et al. Multiple sclerosis patients' benefit-risk preferences: serious adverse event risks versus treatment efficacy. J Neurol. 2009; 256(4):554-62.

64. Bachmann LM, Muhleisen A, Bock A, et al. Vignette studies of medical choice and judgment to study caregivers' medical decision behaviour: systematic review. BMC Med Res Methodol. 2008; 8:50.

65. Johnson FR, Ozdemir S, Mansfield C, et al. Crohn's disease patients' risk-benefit preferences: serious adverse event risks versus treatment efficacy. Gastroenterology. 2007; 133(3):769-79.

66. Tugwell P, Boers M, Strand V, et al. OMERACT 9 - 9th International Consensus Conference on outcome measures in rheumatology clinical trials. J Rheumatol. 2009; 36(8):1765-8.

67. Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. PLoS Med 2011; 8(5):e1001026.

68. Levitan B. A concise display of multiple end points for benefit-risk assessment. Clin Pharmacol Ther. 2011; 89(1):56-9.

69. Sackett DL, Richardson WS, Rosenberg W, et al. Evidence-based medicine. How to practice and teach EBM. Churchill Livingstone, Edinburgh: 1997.

70. JAMA Evidence Glossary [Web Page]. Available at http://jamaevidence.com/glossary/V . Accessed October 21, 2012.

71. Krahn M, Naglie G. The next step in guideline development: incorporating patient preferences. JAMA 2008; 300(4):436-8.

72. Institute of Medicine. Committee on Conflict of Interest in Medical Research, Education, and Practice. Lo B, Field MJ, eds. Conflict of interest in medical research, education, and practice. Washington, DC: National Academies Press; 2009.

73. Guyatt G, Cook D, Haynes B. Evidence based medicine has come a long way. BMJ. 2004; 329(7473):990-1.

74. Guyatt G, Montori V, Devereaux PJ, et al. Patients at the centre: in our practice, and in our use of language. Evid Based Med. 2004; 9(1):6-7.

75. D'Agostino RBJ. Debate: the slippery slope of surrogate outcomes. Curr Control Trials Cardiovasc Med. 2000; 1(2):76-8.

76. Singh S, Loke YK, Furberg CD. Long-term risk of cardiovascular events with rosiglitazone: a meta-analysis. JAMA. 2007; 298(10):1189-95.

77. Loke YK, Price D, Herxheimer A. Systematic reviews of adverse effects: framework for a structured approach. BMC Med Res Methodol. 2007; 7:32.

78. Butler M, Tally KMC, Burns R, et al. Values of Older Adults Related to Primary and Secondary Prevention. Evidence Synthesis No. 84. AHRQ Publication No. 11-05154-EF-1. Rockville, MD: Agency for Healthcare Research and Quality; March 2011.

79. Guyatt GH, Oxman AD, Kunz R, et al. Going from evidence to recommendations. BMJ 2008; 336(7652):1049-51.

80. U.S. Preventive Services Task Force. Methods and Processes. Available at: http://www.uspreventiveservicestaskforce.org/methods.htm .Accessed October 21, 2012.

81. Guyatt G, Akl EA, Hirsh J, et al. The vexing problem of guidelines and conflict of interest: a potential solution. Ann Intern Med. 2010; 152(11):738-41.

82. Shamliyan T, Wyman JF, Ramakrishnan R, et al. Benefits and harms of pharmacologic treatment for urinary incontinence in womena systematic review. Ann Intern Med. 2012 Jun;156(12):861-874. Available at: http://www.annals.org/content/early/2012/04/09/0003-4819-156-12-201206190-00436.full#fn-group-1. Accessed October 21, 2012.

83. Protheroe J, Fahey T, Montgomery AA, et al. Effects of patients' preferences on the treatment of atrial fibrillation: observational study of patient-based decision analysis. West J Med. 2001; 174(5):311-5.

84. Akl EA, Schunemann HJ, Grant B, et al. Working-group for evidence based decision making (WEB Decision Making). Making decisions about using inhaled steroids in COPD. Available at: http://www.predictonline.com/copd/smbs/spm_copd/. Accessed October 21, 2012.

85. Wennberg JE. Tracking medicine: a researcher's quest to understand health care. New York:Oxford University Press; 2010.

86. Fisher ES, Wennberg JE. Health care quality, geographic variations, and the challenge of supply-sensitive care. Perspect Biol Med 2003; 46(1):69-79.

87. Sussman JB, Vijan S, Choi H, et al. Individual and population benefits of daily aspirin therapy: a proposal for personalizing national guidelines. Circ Cardiovasc Qual Outcomes. 2011; 4(3):268-75.

88. Politi MC, Han PK, Col NF. Communicating the uncertainty of harms and benefits of medical interventions. Med Decis Making. 2007; 27(5):681-95.

89. Montori VM, Permanyer-Miralda G, Ferreira-Gonzalez I, et al. Validity of composite end points in clinical trials. BMJ. 2005; 330(7491):594-6.

90. Song F, Loke Y, Walsh T, et al. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ. 2009; 338:b1147 .

Acronyms/Abbreviations

A	Absolute
ARR	Absolute risk reduction
AHRQ	Agency for Healthcare Research and Quality
AHP	Analytic hierarchy process
B&H	Benefit harm
BLRA	Benefit-less-risk analysis
D	Difference
DD	Data driven
EBM	Evidence-based medicine
EPC	Evidence-based Practice Centers
GRADE	Grading of Recommendations, Assessment, Development and Evaluation
INHB	Incremental net health benefit
JAMA	Journal of the American Medical Association
M	Multiple
MAR	Maximum acceptable risk
MCDA	Multicriteria decision analysis
MCE	Minimum clinical efficacy
MERT	Minimum target event risk for treatment
NA	Not applicable
NCB	Net clinical benefit
NCI	National Cancer Institute
NNH	Number needed to harm
NNT	Number needed to treat
NNT/NNH	Ratio of number needed to treat over number needed to harm
NNTt	Threshold number needed to treat
O	Other
OR	Odds ratio
P	Possible
PICOTS	Population, Intervention, Comparator(s) Outcome(s), Timeframe, and Setting
PSM	Probabilistic simulation method
Q-Twist	(Quality-adjusted) time without symptoms and toxicity
QALYs	Quality-adjusted life-years
QFRBA	Quantitative framework for risk and benefit assessment
RBP	Risk-benefit plane
RBC	Risk-benefit contour
RCTs	Randomized controlled trials
RR	Relative Risk
RV-NNT	Relative value adjusted number needed to treat
S	Simulation
SPM	Stated preference method

TURBO	transparent uniform risk benefit overview
Q-TWiST	Time without symptoms and toxicity