BMJ Open

Defining Certainty of Net Benefit: a GRADE concept paper

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-027445.R1
Article Type:	Communication article
Date Submitted by the Author:	26-Oct-2018
Complete List of Authors:	Alper, Brian; EBSCO Health, DynaMed Plus; University of Missouri Columbia School of Medicine, Family and Community Medicine Oettgen, Peter; EBSCO Health, DynaMed Kunnamo, Ilkka; Duodecim Medical Publications Ltd.; The Finnish Medical Society Iorio, Alfonso; McMaster University, Health Research Methods, Evidence, and Impact Ansari, Mohammed; School of Epidemiology, Public Health and Preventive Medicine, Faculty of Medicine, University of Ottawa Murad, M. Hassan; Mayo Clinic, Evidence-Based Practice Center Meerpohl, Joerg; Institute of Medical Biometry & Medical Informatics, University Medical Centre Freiburg, German Cochrane Centre; University Medical Centre, Paediatric Haematology & Oncology, Centre for Paediatrics & Adolescent Medicine Qaseem, Amir; American College of Physicians, Department of Clinical Policy Hultcrantz, Monica; Karolinska Institute; Statens beredning for medicinsk utvardering Schünemann, Holger; McMaster University, Departments of Health Research Methods, Evidence, and Impact and of Medicine, McMaster University Health Sciences Centre, Hamilton, Canada Guyatt, Gordon; Mcmaster University, Health Research Methods, Evidence, and Impact GRADE Working Group, The; GRADE Working Group
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Communication, Epidemiology, Health policy
Keywords:	evidence-based medicine, decision analysis, evidence synthesis, clinical decision making, guideline development

Defining Certainty of Net Benefit: a GRADE concept paper

GRADE defines Certainty of Net Benefit

Brian S Alper, Peter Oettgen, Ilkka Kunnamo, Alfonso Iorio, Mohammed T Ansari, M Hassan Murad, Joerg J Meerpohl, Amir Qaseem, Monica Hultcrantz, Holger Schünemann, Gordon Guyatt, the GRADE Working Group

Brian S Alper, Vice President of Innovations and Evidence-Based Medicine Development, EBSCO Health, EBSCO Information Services, Inc., Ipswich, Massachusetts, United States; Clinical Research Assistant Professor, Department of Family and Community Medicine, University of Missouri-Columbia School of Medicine

Peter Oettgen, Editor in Chief, DynaMed Plus, EBSCO Health, EBSCO Information Services, Inc., Ipswich, Massachusetts, United States

Ilkka Kunnamo, Research Director, Duodecim Medical Publications Ltd.

Alfonso Iorio, Professor, Health Research Methods, Evidence and Impact, McMaster University

Mohammed T Ansari, Adjunct Professor, School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa

M Hassan Murad, Evidence-based Practice Center, Mayo Clinic

Joerg J Meerpohl, Institute for Evidence in Medicine (for Cochrane Germany Foundation); Medical Center – University of Freiburg

Amir Qaseem, Vice President for Clinical Policy and the Center for Evidence Reviews, American College of Physicians

Monica Hultcrantz, Swedish Agency for Health Technology Assessment and Assessment of Social Services; Karolinska Institutet

Holger Schünemann, Chair, Health Research Methods, Evidence and Impact, McMaster University; Director, Cochrane Canada

Gordon Guyatt, Professor, Health Research Methods, Evidence and Impact, McMaster University

The GRADE Working Group

Correspondence to Brian S. Alper, balper@ebsco.com

Word count: 3383

Abstract

Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology is used to assess and report certainty of evidence and strength of recommendations. This GRADE concept article is not GRADE guidance but introduces certainty of net benefit, defined as the certainty that the balance between desirable and undesirable health effects is favorable. Determining certainty of net benefit requires considering certainty of effect estimates, the expected importance of outcomes and variability in importance, and the interaction of these concepts. Certainty of net harm is the certainty that the net effect is unfavorable. Guideline panels using or testing this approach might limit strong recommendations to actions with a high certainty of net benefit or against actions with a moderate or high certainty of net harm. Recommendations may differ in direction or strength from that suggested by the certainty of net benefit or harm when influenced by cost, equity, acceptability, or feasibility.

Article Summary – Strengths and Limitations

- The GRADE Working Group defines certainty of net benefit as the certainty that the balance between desirable and undesirable health effects is favorable.
- The certainty of net benefit offers a concise way to express the certainty that benefits outweigh harms for a recommendation.
- The certainty of net benefit or harm may complement or replace the current GRADE approach to addressing the certainty of evidence associated with recommendations.
- Strengths of this approach include simplicity and clarity for application, and engagement of multiple contributors and stakeholders over two years through the GRADE Working Group.
- Limitations include absence of direct implementation in guideline development; therefore, this GRADE concept article does not constitute GRADE guidance but introduces certainty of net benefit to stimulate discussion and testing in the context of healthcare decision-making.

Introduction

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group has designed a transparent approach to rating certainty of evidence and grading strength of recommendations (1,2). More than one hundred groups creating systematic reviews, clinical practice and public health guidelines, and health technology assessments have adopted GRADE (1,2). GRADE uses the terms "certainty of evidence" interchangeably with "confidence in evidence" and "quality of evidence". Authors using GRADE make separate ratings of certainty for each patient-important outcome and, in the context of a clinical practice guideline, provide an overall rating based on the lowest certainty of the critical outcomes.

In the context of making recommendations, GRADE specifies that ratings reflect the certainty that the estimates of an effect are adequate to support a particular decision or recommendation (3). Recently, the GRADE Working Group clarified the conceptual basis of certainty ratings, noting that, in both contexts of systematic reviews and guidelines, they represent the certainty that a true effect lies on one side of a specified threshold, or within a specified range (4).

Depending on the thresholds or ranges chosen, it is possible to have high certainty in the evidence for a set of outcomes related to a particular decision, yet uncertainty whether the evidence is adequate to support that decision; this will occur when desirable and undesirable consequences are closely balanced, such as cancer treatments with high certainty in prolonging survival and high certainty in serious toxicity (5,6). It is also possible to have low certainty in evidence for a specific outcome yet make a strong recommendation (high certainty to support a decision). The GRADE Working Group has specified five paradigmatic situations in which such discordant recommendations may be appropriate (5,6). One of these situations is when only low quality evidence exists for a promising intervention in a life-threatening context (e.g. using fresh frozen plasma or vitamin K in a patient receiving warfarin with elevated INR and an intracranial bleed).

The recent GRADE Working Group guidance states that systematic review authors and guideline panelists will ideally specify the threshold or ranges they are using when rating the certainty in evidence (3). The guidance offered non-contextualized (no implicit value judgments) and partially contextualized (some implicit value judgments regarding magnitude of effects) approaches for systematic review authors. The guidance further suggested a fully contextualized approach for clinical practice guidelines in which a guideline panel determines thresholds considering all critical outcomes and their relative importance.

Guideline panels using fully contextualized approaches have faced challenges of balancing feasibility and simplicity with comprehensive simultaneous consideration of all important outcomes. This current GRADE concept article introduces an approach for guideline panels to more directly and explicitly rate their certainty of the balance of benefits and harms. This GRADE concept article is presented to stimulate discussion and does not constitute GRADE guidance.

Expressing Certainty Across the Evidence-to-Decision Framework

GRADE Evidence-to-Decision frameworks explicitly identify the following considerations in determining the direction and strength of recommendations:

- Certainty of evidence (regarding effect estimates for health effects) (2,6,7)
- Relative importance of outcomes (also called values and preferences) (2,7,8)
- Balance of benefits and harms (2,7,9)
- Resource use (cost) (2,7,10)
- Cost-benefit ratio (Are incremental health benefits worth the costs?) (2,7,11)
- Equity (7,11)
- Acceptability (11), and
- Feasibility (11).

Health-related harms include pain or disability but also burdens that lower quality of life. For example, the burden of receiving an intervention that requires being immobile for long periods of time could be considered as a health-related harm. In this article, when we use the phrase "balance of benefits and harms" we refer to the "balance of benefits versus harms and burdens". Other burdens that may be considered more societal in nature may be considered through other criteria in the framework (cost, acceptability, feasibility). Here we will use the term "harms" to refer to "health-related harms and burdens".

Ideally, guideline panels consider all the factors listed above when determining the direction and strength of a recommendation. The process may proceed in progressive steps that consider first benefits and harms to generate certainty in net benefit; then costs to generate certainty in a cost-benefit ratio; then equity, acceptability, and equity to address certainty in a recommendation (Figure 1).

Although it makes decisions more transparent, reporting a guideline panel's certainty for each of these concepts may be overwhelming for guideline users seeking simple explanations of the rationale and certainty for recommendations. Among the concepts for which certainty can be expressed formally, the certainty in balance of benefits and harms (net effect) may be most relevant to patients and clinicians (often the primary target users for guidelines). Additional criteria that may influence a recommendation (cost, cost-benefit ratio, equity, acceptability, feasibility) are more likely to vary across social groups and contexts, and population-based ratings may be of less interest to patients and clinicians working together to make individual health care decisions.

Consistent with the recent clarification of "certainty of evidence" - the certainty that a true effect lies within a specified range or on one side of a specified threshold (3) - one can express the certainty of the net effect (or balance of benefits and harms) in terms of a range or in relation to a threshold. The situation when benefits and harms are perfectly balanced (net benefit or harm = 0) represents a natural threshold for certainty of the net effect. Using this threshold, the certainty of net benefit is the certainty that the overall or net effect lies on the side of benefit. The certainty of net harm is the certainty that the net effect lies on the side of harm.

Expressing the certainty of net benefit for guideline users provides the most direct summary representation of the extent of our confidence that the estimates of effects are adequate to support a particular decision or recommendation. The United States Preventive Services Task Force (USPSTF) has used the term certainty of net benefit in a manner consistent with this conceptual framework (12, 13).

Model for Creating the Net Effect Estimate and Rating Its Certainty

Determining the certainty in the balance of benefits and harms involves generating a net effect estimate (a way of specifying the balance of benefits and harms) and then rating the certainty regarding that net effect in relation to the threshold of net benefit = 0 (Figure 2).

Decision analysis provides a statistical method for generating the net effect estimate. Decision modeling has evolved over the years and sophisticated models include multiple outcomes, the varying times at which each outcome can occur, the relative importance placed in each outcome (often using utilities or quality adjusted life years), and future decisions and resulting outcomes. Guideline panels sometimes use decision analysis to evaluate a chain of possible consequences and decisions to inform their recommendations: the UK National Institute for Health and Care Excellence (NICE) relies heavily on such models. Determining the certainty of evidence emerging from such models is itself a complex matter: A GRADE project group is currently addressing the issue.

For many decisions for which guideline developers, clinicians or patients desire recommendations, however, one need not consider a chain of subsequent decisions. In such cases, one can perform a much simpler decision analysis without requiring participation of a skilled modeler. Simple models can generate confidence intervals for a net effect estimate (a composite of individual effect estimates) given the following assumptions (described further in the Appendix):

1. Effect estimates represent data conforming to normal distributions
2. Effect estimates to be combined are independent and not correlated with each other
3. Effect estimates to be combined can be multiplied by a conversion factor to use a consistent unit of measure

Given that the second assumption is often unlikely to hold, the analyst can perform sensitivity analysis of the net effect estimate to determine robustness to changes in the individual effect estimates, the assumptions of correlation between effect estimates, and the conversion factors. A sensitivity analysis defining the likelihood of the net effect estimate remaining favorable across the range of assumptions determines the certainty of net benefit.

Generation of the Net Effect Estimate

Here we describe the methods for generating the net effect estimate as presented in Figure 2. Algorithm-supported calculators can facilitate combining the importance-adjusted effect estimates (the third step in Figure 2) and classifying the precision (the fourth step). The Appendix provides examples and a link to a free online calculator.

Step 1: Determine the outcomes to be combined

Including both a composite outcome and one or more components of that outcome is problematic. For example, it would be inappropriate to include all-cause mortality and cardiovascular mortality in the same model. One may choose to use only the composite outcome (e.g. all-cause mortality) or to use only the component outcomes (e.g. cardiovascular mortality, cancer mortality, and mortality from causes other than cancer or cardiovascular disease).

If effect estimates are not available in absolute terms (or if effect estimates are being extrapolated to a population with different baseline risks than that used for the absolute effect estimates) then absolute effect estimates may be derived using a combination of relative effect estimates and baseline risk estimates.

Step 2: Determine the quantified relative importance for each outcome

Quantitative estimates of relative importance for each outcome will serve as a conversion factor to use a consistent unit of measure for the net effect estimate. These estimates need to be meaningful as a multiplier or represent a quantitative measure of importance relative to a reference standard. Guideline panels that use a qualitative 9-point rating of importance of outcomes to determine which outcomes to include in systematic reviews or summary of findings tables may find these ratings do not easily translate to quantitative estimates for this purpose.

A simple approach is to select one outcome as a reference outcome and define a relative importance adjustment (i.e., a multiplier) for each other outcome as a modifier to apply to effect estimates. In making individual patient-specific decisions, one could enter the quantitative estimates of relative importance for the individual patient and derive an individualized estimate of net effect. With further development this approach could inform shared decision-making for individual patients.

For groups of patients, one could consider quantitative estimates of relative importance as ranges. In making population-specific recommendations, one could use a range of relative importance estimates considered reasonable to capture most members of the population and check for robustness of estimates of net effect across the range of relative importance. One would then lower the rating of certainty of net benefit if the estimate of net effect crosses to net harm within the range of relative importance. The later discussion of sensitivity analysis for the net effect estimate will address the concepts of ranges and certainties of relative importance.

Methods to determine quantitative estimates of relative importance from a patient perspective include discrete-choice experiments (14), preference-eliciting surveys among patients (15), and systematic reviews of such surveys (16). Determination of relative importance could provide an opportunity for engaging patients as partners in research design, a developing expectation in medical publishing (17). When such evidence is unavailable for the outcomes associated with a recommendation, guideline panels can still explicitly make best guesses of the importance the target population will place on the relevant outcomes. Further discussion of the methods for determining relative importance is beyond the scope of this paper.

If the outcomes to be combined include both continuous measures and dichotomous measures, the assignment of relative importance becomes more complicated and would take additional methods to

reach a shared unit of measure (such as conversion to quality-adjusted life-year estimates). Utilities reported for decision analyses may be convertible to relative importance of outcomes. However, utilities are often reported with a range from 0 (for death or worst outcome) to 1 (for optimal quality of life or best outcome), and relative importance of outcomes functioning as multipliers would not be meaningful if multiplied by 0. Relative importance of outcome estimates equal to 1 minus the utility could convert utilities to meaningful multipliers.

Step 3: Combine the importance-adjusted effect estimates

For each effect estimate, one can multiply the point estimate and confidence intervals (CI) by the relative importance for the outcome, and then present the importance-adjusted effect estimate in positive or negative terms to correspond to benefits or harms in the direction of effect.

Adding together the point estimates for each importance-adjusted effect estimate will provide the point estimate for the net effect. Statistical formulas allow calculation of the 95% CI for the net effect (see Appendix).

Rating the Certainty of Net Benefit

Step 4: Classify the precision of the net effect estimate

Precision becomes meaningful with contextual anchoring. Reporting results with a 3-centimeter range would be overly precise for planning travel by car and unacceptable imprecision for some types of surgery. To express the certainty in the balance of benefits and harms, we need to specify a threshold for a net benefit, then express the certainty that the net effect lies on one side of this threshold.

Guideline panels may specify the threshold of net effect; we suggest using the "zero effect" for simplicity. Guideline panels that formally evaluate cost-effectiveness already use a method to set a value threshold for the quantity of net benefit that is considered worth the cost to achieve it.

If the entire confidence interval does not cross zero, then the precision of the net effect estimate is sufficient to not rate down the certainty of net benefit for imprecision. One must still consider other factors affecting certainty that are more difficult to quantify (risk of bias, inconsistency, indirectness, and publication bias) and the plausible range of relative importance of outcomes before final determination of the certainty of net benefit.

If the confidence interval includes zero effect and thus the range of net effect estimates includes both net benefit and net harm, the guideline panel will rate down the certainty of net benefit. The greater the extent of overlap of the confidence interval with both benefit and harm, the lower the certainty in the net benefit. Table 1 and Figure 3 present initial suggestions for how these judgments may be made.

Table 1. Classification of precision of net effect estimate

Pattern of net effect estimate	Classification	Precision of net effect estimate is consistent with …
Entire confidence interval is beneficial	Net benefit	High certainty of net benefit
Point estimate is beneficial, lower bound of confidence interval is harmful, and point estimate has larger absolute value than lower bound of confidence interval	Likely net benefit	Moderate certainty of net benefit
Point estimate is beneficial, lower bound of confidence interval is harmful, and point estimate has smaller absolute value than lower bound of confidence interval	Possible net benefit	Low certainty of net benefit
Point estimate is close to zero, wide confidence interval*	Possibly no net benefit or harm	Very low certainty of net benefit or harm
Point estimate is close to zero, narrow confidence interval*	Net benefit or harm likely near zero	Moderate certainty of little net benefit or harm
Point estimate is harmful, upper bound of confidence interval is beneficial, and point estimate has smaller absolute value than upper bound of confidence interval	Possible net harm	Low certainty of net harm
Point estimate is harmful, upper bound of confidence interval is beneficial, and point estimate has larger absolute value than upper bound of confidence interval	Likely net harm	Moderate certainty of net harm
Entire confidence interval is harmful	Net harm	High certainty of net harm

* Differentiation of wide vs. narrow confidence intervals could be based on a threshold of minimally important differences.

The calculation for confidence intervals for the net effect estimate includes an assumption that effect estimates being combined are not correlated with each other. If effects are correlated, the accurate confidence intervals would be wider or less precise; if inversely correlated, the accurate confidence intervals would be narrower or more precise. If such accuracy is needed, one could add correlation coefficients to the calculation (see Appendix) or rely on more sophisticated statistical approaches such as bootstrapping (18) or a Bayesian approach to estimate the probability interval (19). The calculation is also based on an assumption that effects on outcomes are independent. For practical use, modest violations of the assumption are unlikely to distort results substantially and may be preferable to less explicit judgment of the balance of benefits and harms.

Step 5: Consider the certainty of effect estimates for outcomes that are critical to the likelihood of net benefit.

One approach to select the outcomes critical to the likelihood of net benefit is to identify the outcomes that could change the classification of the precision of the net effect estimate. Such outcomes are either:

- Outcomes for which removal of the outcome would change the classification of the precision of the net effect estimate.
- Outcomes for which gross underestimation of the outcome is plausible and addition of plausible increases to the effect estimate would change the classification.

Determining the lowest certainty of evidence among critical outcomes requires addressing risk of bias, inconsistency, indirectness, and publication bias for each critical outcome. Imprecision for an individual outcome is not an influencing factor here because it is already accounted for in the net effect estimate.

The lowest of the certainty ratings for critical outcomes and the certainty rating consistent with the precision of the net effect estimate represents the certainty of net benefit. This approach may work in most cases; raters still need, however, to consider the overall framework and determine if limited certainty in single outcomes are sufficient to rate down the overall certainty of net benefit. This is especially so if the upper or lower bounds of the CI for the net effect estimate approximates a zero effect. A 95% CI is used based on convention rather than a theoretical rationale.

Step 6: Consider the range of relative importance for outcomes. Perform a sensitivity analysis to determine the certainty of net benefit across this range.

To enhance feasibility of the approach, efforts to fully consider the range of relative importance for outcomes may be limited to ratings that would otherwise be classified as high certainty of net benefit. In situations in which further assessment is needed to confirm robustness of certainty across the range of relative importance, one can repeat the analyses across a reasonable range of relative importance of outcomes.

The purpose of the sensitivity analysis is to determine if the certainty of net benefit remains high across the range of relative importance estimates. There remains insufficient conceptual development to provide explicit guidance on how to precisely define the range of relative importance for outcomes to use for the sensitivity analysis.

The GRADE Working Group has developed guidance on rating the certainty of relative importance of outcomes (20). If a range of relative importance of outcomes is determined by empirical evidence and that range is considered to have low certainty, it would then be prudent to use a wider range of relative importance of outcomes in a sensitivity analysis.

It may be necessary during the process of the sensitivity analysis of outcome importance to re-evaluate which outcomes are critical to the likelihood of net benefit.

Relating Certainty of Net Benefit to Strength of Recommendation

The certainty of net benefit does not necessarily dictate the strength of recommendation. The evidence-to-decision framework also includes cost, cost-benefit ratio, equity, acceptability, and feasibility as considerations that may modify the strength of recommendation. Panels may choose to focus exclusively on net health effects and not include other elements (e.g. some panels choose not to consider costs, and do not formally consider acceptability, feasibility, and equity).

In situations in which there is a high certainty in effect estimates but uncertainty that the balance of benefits and harms is favorable across the range of patient values and preferences (a situation in which panels will make weak recommendations because fully informed patients are likely to make different decisions), a moderate or low certainty of net benefit provides a clear expression of the rationale for weak recommendations.

High certainty is not necessary, in all cases, for supporting a strong recommendation. *Primum non nocere* ("First, do no harm") is considered one of the principal precepts for ethical decision-making in medicine and pharmacology (21) though it is more properly considered *Primum non net nocere* (22). One can interpret this to consider a lower threshold for the certainty in net harm for a strong recommendation against an action than one uses for the certainty in net benefit for a strong recommendation for an action.

Implications

In this article, we introduce an approach for guideline developers to consider explicitly reporting the certainty of net benefit with recommendations, either in addition to or in place of reporting an overall quality of evidence associated with a recommendation. Either way, the approach requires consideration of certainty of evidence ratings for individual outcomes, typically presented in summary of findings tables.

This approach involves many judgments that are already made explicitly or implicitly when guideline panels make recommendations. Reporting the judgments made when using this approach would allow

readers to interpret their confidence in how the ratings were made and may reduce spurious confidence that could occur with quantitative reporting in the absence of qualitative factors.

A key driver for this approach is greater congruence with the intent behind the concept of "adequate evidence to support a recommendation" than what is currently conveyed by the "overall quality of evidence in estimates of effects". Strengths of this approach include the transparent, logical, quantitative expressions for both scholarly and clinical readers, and both guideline developers and guideline users.

The primary limitation of this approach is its lack of testing to inform its feasibility and acceptability, and how readers will interpret these concepts. This report is shared, before such testing, to increase scholarly discussion. This GRADE concept article does not therefore constitute GRADE guidance.

Acknowledgements: The authors are pleased to acknowledge Lehana Thabane, PhD for collaborating with the lead author to derive the statistical method; Paul E. Alexander, MSc, MHSc, Cynthia Boyd, MD, MPH, Reem Mustafa, MD and Irfan Dhalla MD, MSc for feedback to improve the readability and conceptual clarity of the manuscript; and Jordan Prince for construction of an online calculator.

Contributorship statement: At a GRADE Working Group (GWG) meeting GG and MH presented a complex method to rate the certainty of evidence for an outcome when fully contextualized with respect to other outcomes related to a decision or recommendation. BSA introduced the concept of certainty of net benefit to clarify and simplify methodology to report and assess the balance of benefits and harms in the context of fully contextualizing certainty of evidence across outcomes. The GWG formed a subcommittee for this conceptual development. BSA and Lehana Thabane developed the statistical model for a simple decision analysis and sensitivity analysis. BSA, IK, AI and Lehana Thabane provided examples to demonstrate the model (Appendix). The GWG had in-person meetings in three countries (with up to 100 people in attendance) in which the wider audience provided in-depth review, feedback and discussion to refine the concepts. BSA, PO, IK, AI, MTA, MHM, JJM, AQ, MH, HS and GG met frequently to iteratively refine the concepts, meet the authorship requirements, and approve the final version. BSA is the guarantor of the article.

Funding: No funding support.

Competing interests: All authors have completed the unified competing interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author). There are no direct competing interests as the article provides conceptual development and no promotion or assessment of any product or service. All authors are members of the GRADE Working Group and conduct scholarly activity or professional services related to the concepts in this article. BSA and PO are employed by EBSCO Information Services and IK is employed by Duodecim Medical Publications Ltd.

Transparency: BSA affirms that that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as originally planned (and, if relevant, registered) have been explained.

Data Sharing Statement: No additional data available

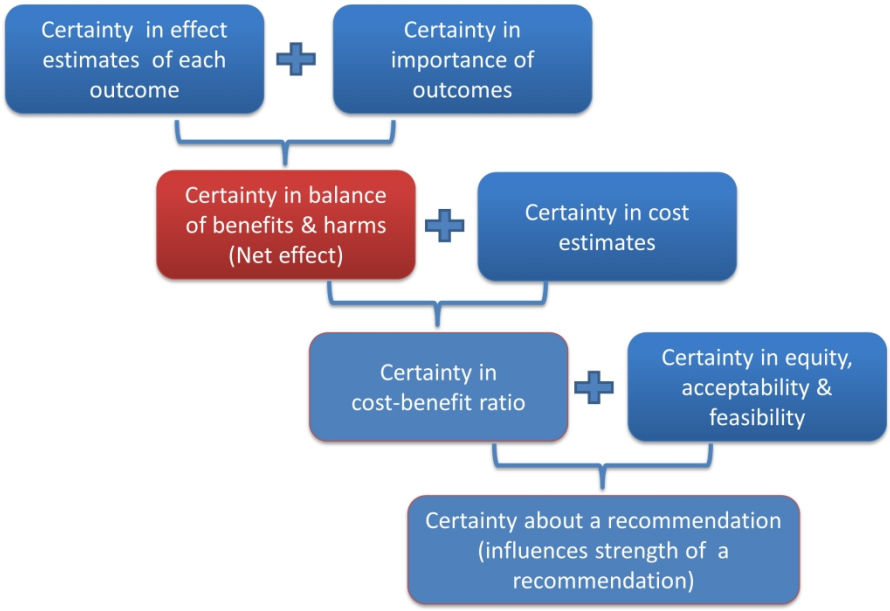
Patient and Public Involvement statement: The manuscript was shared with the Patient Editor for Research and Evaluation at The BMJ (Amy Price) who welcomed the development and suggested Patient

Involvement was not directly appropriate for this report. Public Involvement was not openly pursued but the GRADE Working Group provided multiple opportunities for more than 100 multidisciplinary stakeholders to contribute.

References:

1. URL: www.gradeworkinggroup.org. Accessed November 27, 2016.
2. Atkins D, Best D, Briss PA, et al; GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ 2004;328:1490. doi: 10.1136/bmj.328.7454.1490.
3. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;64:401-6. doi: 10.1016/j.jclinepi.2010.07.015.
4. Hultcrantz M, Rind D, Akl E, et al; GRADE Working Group. The GRADE Working Group clarifies the construction of certainty of evidence. J Clin Epidemiol 2017;87:4-13. doi: 10.1016/j.jclinepi.2017.05.006.
5. Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol 2013;66:726-35. doi: 10.1016/j.jclinepi.2013.02.003.
6. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.3.2 Confidence in best estimates of magnitude of effects (quality of evidence). URL: gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.s1fsemn2yjpm. Accessed November 27, 2016.
7. Guyatt GH, Oxman AD, Kunz R, et al; GRADE Working Group. Going from evidence to recommendations. BMJ 2008;336(7652):1049-51. doi: 10.1136/bmj.39493.646875.AE.
8. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.3.3 Confidence in values and preferences. URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.i5hfweocv3qs. Accessed November 27, 2016.
9. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.2.1 Balance of desirable and undesirable consequences. URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.feb5g22tf6nv. Accessed November 27, 2016.
10. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.3.4 Resource use (cost). URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.y0ey8465h47v. Accessed November 27, 2016.
11. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.5 The Evidence-to-Decision framework. URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.xr5ac2p2khuq. Accessed November 27, 2016.
12. Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D, U.S. Preventive Services Task Force. Update on the methods of the U.S. Preventive Services Task Force: estimating certainty and magnitude of net benefit. Ann Intern Med 2007;147:871-5. doi: 10.7326/0003-4819-147-12-200712180-00007.

13. Krist AH, Wolff TA, Jonas DE, et al. Update on the methods of the U.S. Preventive Services Task Force: methods for understanding certainty and net benefit when making recommendations. Am J Prev Med 2018;54:S11-S18. doi: 10.1016/j.amepre.2017.09.011.
14. Hazlewood GS, Bombardier C, Tomlinson G, Marshall D. A Bayesian model that jointly considers comparative effectiveness research and patients' preferences may help inform GRADE recommendations: an application to rheumatoid arthritis treatment recommendations. J Clin Epidemiol 2018;93:56-65. doi: 10.1016/j.jclinepi.2017.10.003.
15. Fain KM, Yu T, Li T, Boyd CM, Singh S, Puhan MA. Evidence selection for a prescription drug's benefit-harm assessment: challenges and recommendations. J Clin Epidemiol 2016;74:151-7.
16. MacLean S, Mulla S, Akl EA, et al. Patient Values and Preferences in Decision Making for Antithrombotic Therapy: A Systematic Review: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 2012;141:e1S-e23S.
17. Richards T, Schroter S, Price A, Godlee F. Better together: patient partnership in medical journals. BMJ 2018;362:k3798.
18. Efron B, Tibshirani RJ. An Introduction to the Bootstrap, 1st Edition Chapman and Hall/CRC. New York, NY 1993.
19. Spiegelhalter DJ. Myles JP, Jones DR, Abrams KR. An introduction to Bayesian methods in health technology assessment. BMJ 1999;319:508–512.
20. Zhang Y, Alonso Coello P, Guyatt G, Yepes-Nunez JJ, Akl EA, Hazlewood G, et al. GRADE Guidelines: 19. Assessing the certainty of evidence in the importance of outcomes or values and preferences -- Risk of bias and indirectness. J Clin Epidemiol 2018.
21. Smith CM. Origin and uses of primum non nocere—above all, do no harm! J Clin Pharmacol 2005;45:371-7. doi: 10.1177/0091270004273680
22. Schunemann HJ. Guidelines 2.0: do no net harm-the future of practice guideline development in asthma and other diseases. Current allergy and asthma reports 2011;11:261-8.



* This figure is not intended to convey a hierarchy, order or requirement of factors for consideration. The red component represents the key concept of this discussion.

Figure 1. Certainty across the evidence-to-decision framework*

254x190mm (300 x 300 DPI)

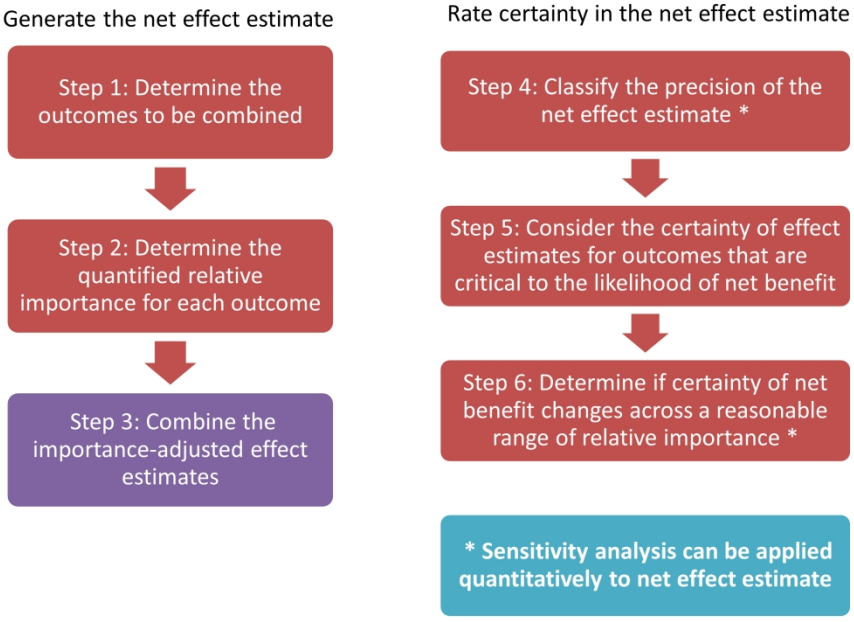


Figure 2. A stepwise approach to determining the certainty of the net effect estimate

254x190mm (300 x 300 DPI)

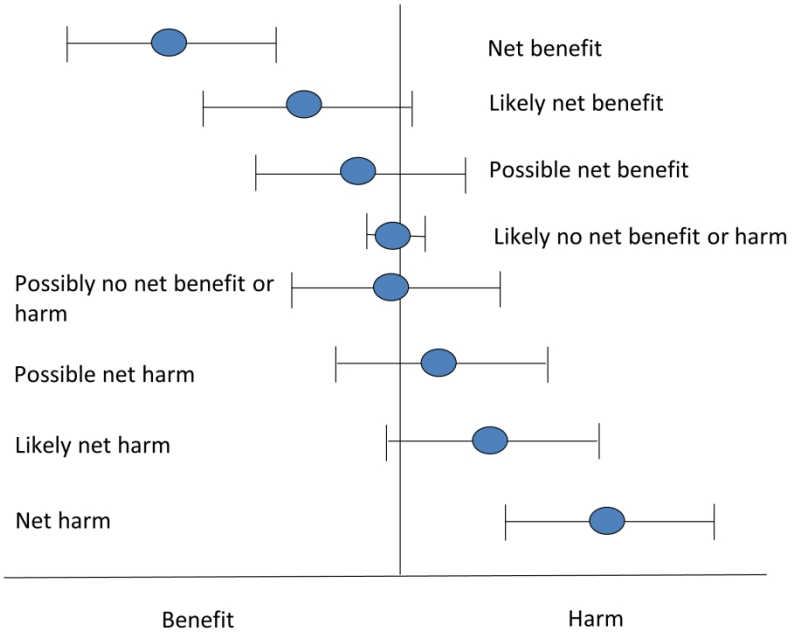


Figure 3. Classification of precision of net effect estimate

254x190mm (300 x 300 DPI)

Supplemental Material (Appendix)

Calculating the Net Effect Estimate

The statistical method is fully described here and a free online calculator is available at ebscohealth.com/innovations. In brief, while we add the point estimates for each effect estimate to determine the point estimate for the net effect estimate, the calculation of the 95% confidence interval requires a couple of formulas.

Suppose we have an effect estimate X for one outcome and Y for another outcome and we want to determine a combined or net effect estimate Z. The model to determine the net effect estimate Z as a summative or linear combination of effect estimate X and effect estimate Y is based on the same statistical principles for determination of confidence intervals for differences between means, using addition instead of subtraction. That is,

$$Z = X + Y.$$

Assumptions regarding effect estimates include they:

1) represent data conforming to the normal distribution,
2) are independent and not correlated with each other, and
3) are expressed using the same units of measure.
4) The mean (or point estimate) for a net effect estimate Z is simply the addition of the means (or point estimates) for effect estimates X and Y. That is,

$$\text{Mean } Z = \text{Mean } X + \text{Mean } Y.$$

A 95% confidence interval for the net effect estimate is determined by calculating Mean Z +/- 1.96 SD_{MeanZ} where SD_{MeanZ} = standard deviation [SD] of Mean Z.

For the net effect estimate Z, the SD of Mean Z is related to the SDs of the component estimates Mean X and Mean Y through the formula:

$$SD_{MeanZ}^2 = SD_{MeanX}^2 + SD_{MeanY}^2$$

Therefore,

$$SD_{MeanZ} = \sqrt{(SD_{MeanX}^2 + SD_{MeanY}^2)}$$

The 95% confidence interval for the net effect will be:

$$\text{Mean } Z - 1.96\ SD_{MeanZ} \text{ to Mean } Z + 1.96\ SD_{MeanZ}$$

The third assumption (that effect estimates X and Y are expressed using the same units of measure) is rarely true so we need to introduce a “standardization” or “normalization” of outcomes, and this can be

done based on their relative importance. One approach is to assign a multiplier (M) to each outcome representing its importance or relative value compared to a reference outcome. The reference outcome can be external to the body of evidence, or can be one of the outcomes of interest (in which case the value of M for the reference outcome will be 1).

The mean (or point estimate) for a net effect estimate Z, expressed in units of multiples of the reference outcome, becomes:

$$\text{Mean } Z = (M_X \times \text{Mean } X) + (M_Y \times \text{Mean } Y)$$

With the use of multipliers, the SD of the net effect estimate Z becomes related to the formula:

$$SD_{MeanZ}^2 = (M_X \times SD_{MeanX})^2 + (M_Y \times SD_{MeanY})^2$$

Therefore,

$$SD_{MeanZ} = \sqrt{(M_X^2 SD_{MeanX}^2 + M_Y^2 SD_{MeanY}^2)}$$

Note that if the SD_{Mean} is not directly reported for an individual effect estimate, it can be derived from the width of the 95% confidence interval (CIW) for the effect estimate:

$$SD_{MeanX} = CIW_X / 3.92$$

Using the data for the sacubitril-valsartan example (with units of hospitalization-equivalent events per 1000 patients) we get:

$SD_{\text{All-cause mortality outcome}}$ = CIW of 160 / 3.92 = 40.816

$SD_{\text{Hospitalization rate outcome}}$ = CIW of 27 / 3.92 = 6.888

$SD_{\text{Symptomatic hypotension rate outcome}}$ = CIW of 12 / 3.92 = 3.061

Applying $SD_{MeanZ} = \sqrt{(SD_{MeanX}^2 + SD_{MeanY}^2)}$ we get:

$SD_{\text{Net effect estimate}} = \sqrt{(SD_{\text{All-cause mortality}}^2 + SD_{\text{Hospitalization rate}}^2 + SD_{\text{Symptomatic hypotension rate}}^2)}$

$SD_{\text{Net effect estimate}} = \sqrt{(40.816^2 + 6.888^2 + 3.061^2)}$

$SD_{\text{Net effect estimate}} = \sqrt{(1665.9459 + 47.4445 + 9.3697)}$

$SD_{\text{Net effect estimate}} = \sqrt{(1722.7601)} = 41.5$

The 95% confidence interval for the net effect estimate is the mean +- 1.96 SD. For the lower boundary, this translates to 154 - (1.96)(41.5) = 154-81.34 = 72.66 (rounded to 73) and for the upper boundary, this would be 154 + 81.34 = 235.34 (rounded to 235).

We report a net effect estimate of a decrease in 154 hospitalization-equivalent events per 1000 patients (95% confidence interval for the net effect estimate being 73 fewer to 235 fewer hospitalization-equivalent events per 1000 patients).

Sensitivity analysis of the net effect estimate

The 95% confidence interval implies a range within which there is 95% certainty that the true net effect occurs. There are many factors that can affect the certainty that the true net effect is within this range (or affect the precision of the range for which there is 95% certainty of including the true net effect).

If assumptions used in the model are not met, the results will not have accurate precision. If individual outcomes are correlated (such as increase in one benefit being correlated with an increase in another benefit), the "true" 95% confidence interval would be wider or less precise than the one estimated by our method. Alternatively, if individual outcomes are inversely correlated (such as an increase in a benefit being correlated with an increase in a harm, or correlated with a decrease in another benefit), then the "true" 95% confidence interval would be narrower or more precise than the one estimated by our method. In the latter case our proposed approach is conservative but less powerful. If outcomes have other dependencies or do not follow a normal distribution (such as a highly skewed distribution), then the 95% confidence interval may be inaccurate.

The statistical formulas can be adjusted with correlation coefficients if they can be estimated. The formula to determine the standard deviation of the mean of the net effect

$$SD_{MeanZ}^2 = (M_X \times SD_{MeanX})^2 + (M_Y \times SD_{MeanY})^2$$

is modified to

$$SD_{MeanZ}^2 = (M_X \times SD_{MeanX})^2 + (M_Y \times SD_{MeanY})^2 + (2 \times r \times M_X \times SD_{MeanX} \times M_Y \times SD_{MeanY})$$

where r = the correlation coefficient between X and Y. Correlation coefficients are rarely available but the maximum value of r that appears plausible can be used for a sensitivity analysis to address plausible correlations between outcomes.

For the sacubitril-valsartan example, there is data suggesting a small inverse correlation (r = -0.17) between all-cause mortality and hospitalization for heart failure among patients with heart failure (43). There is no data addressing correlations between drug-related symptomatic hypotension and the outcomes of mortality or hospitalization. Let's assume r = 0.5 for each of these as an upper bound of plausible correlations for a sensitivity analysis.

$$SD_{\text{Net effect estimate}} = \sqrt{(SD_{\text{All-cause mortality}}^2 + SD_{\text{Hospitalization rate}}^2 + SD_{\text{Symptomatic hypotension rate}}^2) + 2r(SD_{\text{Mortalty}})(SD_{\text{Hospiltalization}}) + 2r(SD_{\text{Mortalty}})(SD_{\text{Symptomatic hypotension rate}}) + 2r(SD_{\text{Symptomatic hypotension rate}})(SD_{\text{Hospitalization}})}$$

$$SD_{\text{Net effect estimate}} = \sqrt{(40.816^2 + 6.888^2 + 3.061^2) + 2(-0.17)(40.816)(6.888) + 2(0.5)(40.816)(3.061) + 2(0.5)(3.061)(6.888)}$$

$$SD_{\text{Net effect estimate}} = \sqrt{(1722.7601) + (-95.5878) + (124.9378) + (21.0842)}$$

$$SD_{\text{Net effect estimate}} = \sqrt{(1773.1943)} = 42.1$$

This net effect estimate (in a sensitivity analysis adjusting for known and plausible correlations among outcomes) is a decrease in 154 hospitalization-equivalent events per 1000 patients (95% confidence interval 71 fewer to 237 fewer hospitalization.

Example 1. Longer dual-antiplatelet therapy (DAPT) after drug-eluting stents

A systematic review comparing longer versus shorter durations of DAPT after drug eluting stent placement provides the summary of effect estimates for longer duration DAPT in Appendix Table 1 (23). Longer duration of DAPT ranged from 12 months to 42 months and shorter duration of DAPT ranged from 3 months to 18 months (23).

Appendix Table 1. Summary of findings for longer versus shorter durations of dual antiplatelet therapy after drug eluting stent placement

Outcome	Absolute effect estimate per 1000 patients (95% confidence interval)	Certainty of effect estimates
All-cause mortality	2 more (0 change to 4 more)	High*
Myocardial infarction	8 fewer (12 fewer to 2 fewer)	Moderate
Major bleeding	6 more (3 more to 10 more)	High
Any stroke	0 change (2 fewer to 2 more)	High*

* originally reported as moderate quality evidence with downgrade limited to precision. Precision downgrade for single outcome effect estimates are not relevant in this approach as the confidence intervals are being used in the determination of the net effect estimate.

Step 1. Determine the outcomes to be combined.

All four outcomes (mortality, myocardial infarction, major bleeding, and stroke) are considered impactful to include in net effect estimates. None are overlapping outcomes with the assumptions that hemorrhagic stroke contributes minimally to estimates of major bleeding, and fatal outcomes contribute minimally to estimates of myocardial infarction, major bleeding and stroke.

Step 2. Determine the quantified relative importance for each outcome.

Myocardial infarction-equivalent will be considered the reference unit. For example purposes, we will start with the assumption that patients would consider the importance of a myocardial infarction and major bleeding similarly, consider a stroke 3 times more important, and consider mortality 5 times more important. These assignments of relative importance of outcomes are derived from systematic review of evidence of relative importance of outcomes for myocardial infarction, major bleeding and stroke (16) and without empiric investigation for the mortality outcome (3).

Step 3. Combine the importance-adjusted effect estimates.

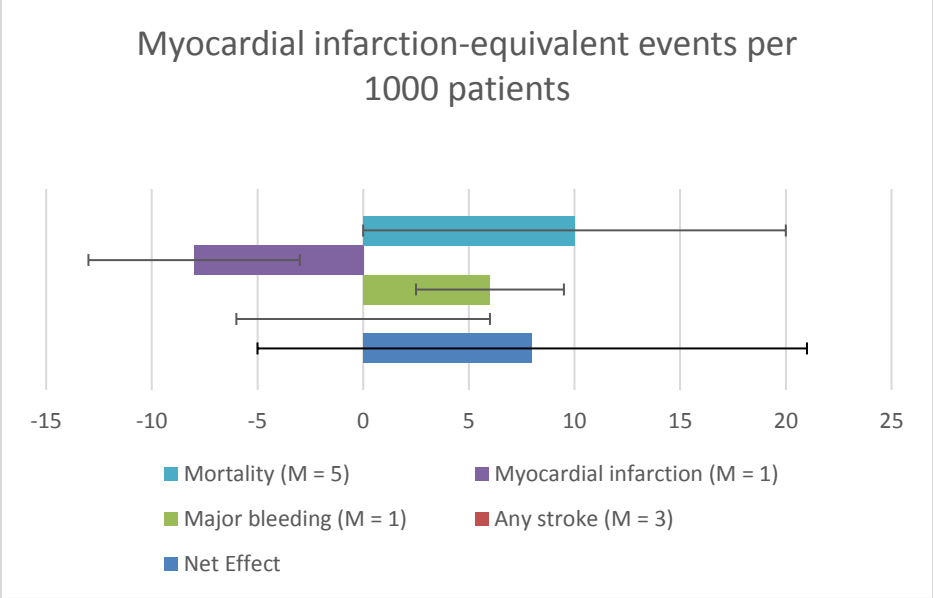
Importance-adjusted effect estimates are determined by multiplying each effect estimate by its relative importance multiplier. Our importance-adjusted effect estimates (in units of myocardial infarction-equivalent events per 1000 patients) are summarized in Appendix Table 2:

Appendix Table 2. Importance-adjusted effect estimates for longer versus shorter durations of dual antiplatelet therapy after drug eluting stent placement

Outcome	Absolute effect estimate (in units of myocardial infarction-equivalent events per 1000 patients) (95% confidence interval)
All-cause mortality	10 more (0 change to 20 more)
Myocardial infarction	8 fewer (12 fewer to 2 fewer)
Major bleeding	6 more (3 more to 10 more)
Any stroke	0 change (6 fewer to 6 more)

The effect estimates are combined using the online calculator at ebscohealth.com/innovations (see Appendix Part 2). The net effect estimate is an increase in 8 myocardial infarction-equivalent events per 1000 patients (95% confidence interval [CI] decrease in 5 to increase in 21 myocardial infarction-equivalent events per 1000 patients).

Appendix Figure 1. Effect estimates for longer versus shorter durations of dual antiplatelet therapy after drug eluting stent placement



Step 4. Classify the precision of the net effect estimate.

The net effect point estimate is harmful, the lower bound of the confidence interval for the net effect estimate is beneficial, and the absolute value of the lower bound of the confidence interval is smaller than the absolute value of the net effect point estimate. This pattern is likely net harm, and consistent with a moderate certainty of net harm.

Step 5. Consider the certainty of effect estimates for outcomes that are critical to the likelihood of net benefit.

Mortality and major bleeding are critical outcomes (potential differentiators of the likelihood of net benefit) because removal of either outcome could change the pattern to one suggesting net benefit. Stroke and myocardial infarction have limited impact on the net effect classification. Both critical outcomes have high certainty of evidence so this does not change our moderate certainty of net harm.

Step 6. Consider the range of relative importance for outcomes. Determine if the net effect estimate across the range of relative importance changes the certainty of net benefit rating.

If patients considered reduction of myocardial infarction to have higher relative importance than mortality and major bleeding it is possible to derive a net benefit. Such relative importance ratings are plausible because myocardial infarction can have a greater contribution to long-term quality of life. Consideration of the range of relative importance for outcomes leads to a low certainty of net harm.

Completing the evidence-to-decision framework

With a low certainty of net harm, the expected result is a weak recommendation against longer duration DAPT after drug-eluting stent placement. The costs are relatively low and there are little adverse consequences related to acceptability, feasibility and equity, so guideline panels may consider to make a weak recommendation against longer duration DAPT.

At the current time, major guidelines have inconsistent recommendations for this concept. The American College of Chest Physicians makes a strong recommendation against DAPT (and for single antiplatelet therapy) after 12 months following drug-eluting stent placement (24). The American College of Cardiology makes a weak recommendation suggesting continuing DAPT beyond 12 months may be considered in patients receiving drug-eluting stents (25).

Example 2. Sacubitril-valsartan for symptomatic heart failure

This example is a decision or recommendation to use sacubitril-valsartan instead of an angiotensin-converting enzyme (ACE) inhibitor or angiotensin receptor blocker (ARB) in patients with symptomatic heart failure with reduced ejection fraction despite treatment with an ACE inhibitor or ARB. A systematic evidence review and GRADE evidence profile for such use of sacubitril-valsartan finds the effect estimates in Appendix Table 3 (26,27), based on a single trial (28).

Appendix Table 3. Summary of findings for sacubitril-valsartan versus continued ACE inhibitor or ARB in symptomatic heart failure with reduced ejection fraction

Outcome	Absolute effect estimate per 1000 patients (95% confidence interval)	Certainty of evidence*
All-cause mortality	29 fewer (12 fewer to 44 fewer)	Moderate
Cardiovascular mortality	31 fewer (17 fewer to 45 fewer)	Moderate
Hospitalization for worsening heart failure	31 fewer (16 fewer to 43 fewer)	Moderate
Symptomatic hypotension	44 more (33 more to 57 more)	Moderate
Change in heart failure symptom score (scale 0-100)	1.64 points decrease (0.63-point decrease to 2.65-point decrease)	Moderate
Decline in renal function	4 fewer (3 fewer to 9 fewer)	Moderate

* Certainty of evidence ratings here do not rate down for imprecision.

Step 1. Determine the outcomes to be combined

In the sacubitril-valsartan example decline in renal function was considered not impactful for determination of the net effect estimate because the effect size is small and the outcome has low importance to patients. Change in heart failure symptom score was considered not impactful for determination of the net effect estimate because the effect size is small and the relative importance is uncertain and may be accounted for in other outcomes. Using means for a continuous score can be misleading when one considers the impact on individual patients who vary in their responses (i.e. assuming every patient experiences the mean effect is likely an erroneous assumption). The only data regarding the proportion of patients who have an important change in symptoms is the outcome of hospitalization for worsening heart failure, and authors of the study reported this outcome.

All-cause mortality is selected instead of cardiovascular mortality to avoid duplicate counting of mortality. The outcomes included in net effect estimation are all-cause mortality, hospitalization for worsening heart failure and symptomatic hypotension.

Step 2. Determine the quantified relative importance for each outcome

Hospitalization-equivalent events per 1000 patients will be considered the reference unit. We do not readily find empiric evidence for the relative importance of outcomes in patients with heart failure. We will start with the assumption that patients would consider the outcome of all-cause mortality 5 times

more important than an episode of hospitalization, and an outcome of symptomatic hypotension half as important as being hospitalized.

Step 3. Combine the importance-adjusted effect estimates

Importance-adjusted effect estimates are determined by multiplying each effect estimate by its relative importance multiplier. Our importance-adjusted effect estimates (in units of hospitalization-equivalent events per 1000 patients) are summarized in Appendix Table 4:

Appendix Table 4. Importance-adjusted effect estimates for sacubitril-valsartan versus continued ACE inhibitor or ARB in symptomatic heart failure with reduced ejection

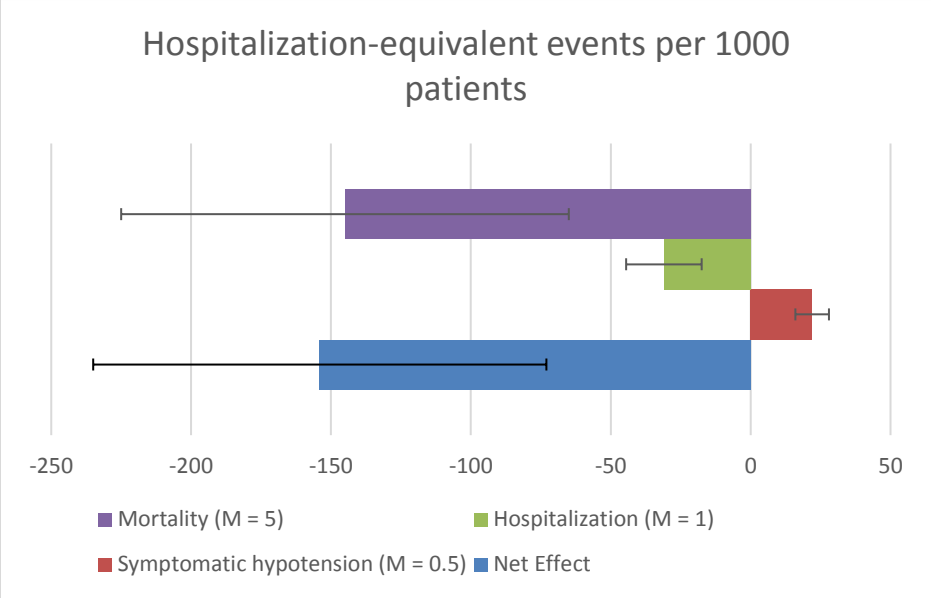
Outcome	Absolute effect estimate (in units of hospitalization-equivalent events per 1000 patients) (95% confidence interval)
All-cause mortality	145 fewer (60 fewer to 220 fewer)
Hospitalization for heart failure	31 fewer (16 fewer to 43 fewer)
Symptomatic hypotension	22 more (16.5 more to 28.5 more)

The effect estimates are combined using the online calculator at ebscohealth.com/innovations and the calculations are shown in part in Appendix Part 2.

The net effect point estimate is a decrease in 154 hospitalization-equivalent events per 1000 patients. (-145 plus -31 plus +22 = -154)

The net effect estimate is a decrease in 154 hospitalization-equivalent events per 1000 patients. (95% CI 73 fewer to 235 fewer hospitalization-equivalent events per 1000 patients)

Appendix Figure 2. Effect estimates for sacubitril-valsartan versus continued ACE inhibitor or ARB in symptomatic heart failure with reduced ejection fraction



Step 4. Classify the precision of the net effect estimate.

The entire confidence interval of the net effect estimate is beneficial so the pattern is net benefit, consistent with a high certainty of net benefit.

Step 5. Consider the certainty of effect estimates for outcomes that are critical to the likelihood of net benefit.

Mortality is potentially differentiating because removal of a mortality effect would change the net effect estimate from 154 fewer (95% CI 73 fewer to 235 fewer) hospitalization-equivalent events per 1000 patients to 9 fewer (95% CI 24 fewer to 6 more) events per 1000 patients, and the overall pattern would change from net benefit to likely net benefit.

Hospitalization for heart failure is not potentially differentiating because removal from the net effect estimate would not change the pattern from net benefit. The net effect estimate would be 123 fewer (95% CI 43 fewer to 203 fewer) events per 1000 patients. A result of increasing hospitalizations for heart failure is not a plausible likelihood. One could question whether total hospitalizations should be used as an outcome rather than cause-specific hospitalization. The outcome of total hospitalizations was not reported in the underlying evidence (28), and guideline panels would need to determine if such an outcome is impactful enough to reassess the overall balance of benefits and harms for this decision.

Symptomatic hypotension is initially not potentially differentiating because removal from the net effect estimate would not change the pattern from net benefit. Symptomatic hypotension can still be considered critical because a higher rate of symptomatic hypotension than observed in the underlying

evidence is plausible, especially related to the use of run-in periods excluding patients who did not tolerate study medications.

The critical outcomes have effect estimates with moderate certainty. This leads to a moderate certainty of net benefit.

Step 6. Consider the range of relative importance for outcomes. Determine if the net effect estimate across the range of relative importance changes the certainty of net benefit rating.

We started with an assumption that the average patient would consider the importance of all-cause mortality five times more important than an episode of hospitalization, and an outcome of symptomatic hypotension half as important as being hospitalized. To consider a range of relative importance for outcomes we should consider the lowest relative importance for all-cause mortality and highest relative importance for symptomatic hypotension that would occur among patients facing this decision and is considered reasonable to reflect the range of importance among common, rational people. Some patients (such as those with terminal illness) may place higher importance on how they feel than mortality so for these patients they might consider mortality, symptomatic hypotension, and hospitalization to be equivalent.

Using assumptions of equivalence across these three outcomes the net effect estimate would be 16 fewer (95% CI 40 fewer to 9 more) events per 1000 patients.

With a reasonable limit for the range of relative importance (including most patients) weighted to support net harm, the net effect estimate changes from net benefit to likely net benefit. If there were otherwise high certainty of net benefit this finding could reduce our certainty to moderate certainty of net benefit. As we already have a moderate certainty of net benefit, extreme assumptions reaching likely net benefit do not further change our certainty.

Completing the evidence-to-decision framework

In an assessment in 2015 the moderate certainty of net benefit justified a weak recommendation for sacubitril-valsartan (26, 27). The high cost further supported a weak recommendation. Four national guidelines have since made strong recommendations for the use of sacubitril-valsartan (29-32), though the findings have not been replicated in a second trial. A recommendations panel reconsidered the rationale across all four guidelines and reconfirmed a weak recommendation for sacubitril-valsartan based on a moderate certainty of evidence (limited to a single trial with potential selection bias related to the run-in period), a moderate certainty of net benefit (considering the range of quantitative estimates of importance of outcomes), and high cost with some uncertainty in the cost-benefit ratio (26). A different recommendations panel could generate different ratings, but the process allows explicit and transparent expression of what is being rated and how it is rated.

Example 3. Ivabradine for symptomatic heart failure

Ivabradine is a heart rate lowering drug which has been tested for clinical use in patients with heart failure in two large randomized trials (33, 34). In the first trial ivabradine was not associated with overall clinical benefit and was not associated with any decrease in death or hospitalization attributed to heart failure (33). In the second trial with more stringent selection criteria (left ventricular ejection fraction <= 35%, heart rate => 70 beats/minute) ivabradine reduced the rate of hospital admissions for worsening heart failure (34).

Outcome differences with ivabradine instead of placebo (from randomization until first event, up to 42 months) are summarized in Appendix Table 5 (35):

Appendix Table 5. Summary of findings for ivabradine versus placebo in symptomatic heart failure with reduced ejection fraction

Outcome	Absolute effect estimate per 1000 patients (95% confidence interval)	Certainty of effect estimates
All-cause mortality	13.9 fewer (31.8 fewer to 4 more)	Moderate
Cardiovascular mortality	11.9 fewer (29 fewer to 5.2 more)	Moderate
Death from heart failure	11.4 fewer (21 fewer to 1.8 fewer)	Moderate
Hospitalization for any cause	35.6 fewer (59.4 fewer to 11.8 more)	Moderate
Hospitalization for cardiovascular reason	42.3 fewer (65 fewer to 19.6 more)	Moderate
Hospitalization for worsening heart failure	47.3 fewer (66 fewer to 28.6 fewer)	Moderate
Bradycardia	33.5 more (26 more to 41 more)	High
Phosphenes (a visual adverse effect)	22.6 more (16.5 more to 28.8 more)	High
Atrial fibrillation	12.1 more (2.1 more to 22 more)	High

The second trial had a low risk of bias though the quality of evidence could be considered moderate for benefits based on inconsistency with the first trial. The adverse effects data could be considered as high quality evidence as the findings are consistent with the first trial (36).

Step 1. Determine the outcomes to be combined.

All-cause mortality, hospitalization for any cause, bradycardia, phosphenes, and atrial fibrillation are selected as non-overlapping outcomes.

Step 2. Determine the quantified relative importance for each outcome.

Hospitalization-equivalent relative importance will be estimated at 0.3 for each adverse effect and 5 for mortality.

Step 3. Combine the importance-adjusted effect estimates

The importance-adjusted effect estimates (in units of hospitalization-equivalent events per 1000 patients) are in Appendix Table 6.

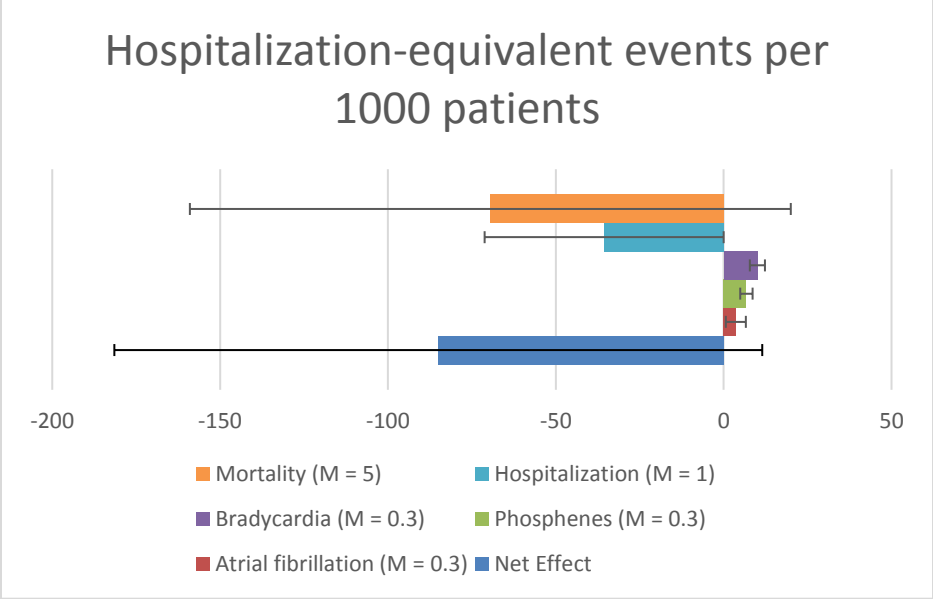
Appendix Table 6. Importance-adjusted effect estimates for ivabradine versus placebo in symptomatic heart failure with reduced ejection fraction

Outcome	**Absolute effect estimate** (in units of hospitalization-equivalent events per 1000 patients) **(95% confidence interval)**	**Confidence interval width (CIW)**
All-cause mortality	69.5 fewer (159 fewer to 20 more)	179 per 1000
Hospitalization for any cause	35.6 fewer (59.4 fewer to 11.8 more)	71.2 per 1000
Bradycardia	10.05 more (7.8 more to 12.3 more)	4.5 per 1000
Phosphenes (a visual adverse effect)	6.78 more (4.95 more to 8.64 more)	3.69 per 1000
Atrial fibrillation	3.63 more (0.63 more to 6.6 more)	5.97 per 1000

The net effect point estimate is a decrease in 85 hospitalization-equivalent events per 1000 patients. (-69.5 plus -35.6 plus +10.05 plus +6.78 plus +3.63 = -84.64)

The net effect estimate is a decrease in 85 (95% confidence interval decrease in 181 to increase in 12) hospitalization-equivalent events per 1000 patients (see Appendix Figure 3).

Appendix Figure 3. Effect estimates for ivabradine versus placebo in symptomatic heart failure with reduced ejection fraction



Step 4. Classify the precision of the net effect estimate.

The pattern is likely net benefit, consistent with a moderate certainty of net benefit.

Step 5. Consider the certainty of effect estimates for outcomes that are critical to the likelihood of net benefit.

As the effect estimates have at least moderate certainty of evidence there is moderate certainty of net benefit.

Step 6. Consider the range of relative importance for outcomes. Determine if the net effect estimate across the range of relative importance changes the certainty of net benefit rating.

A plausible range of relative importance could include consideration of death equivalent to hospitalization and other adverse effects 0.6 times as disruptive as hospitalization. Such assumptions would lead to importance-adjusted effect estimates of:

- All-cause mortality: 13.9 fewer (95% CI 31.8 fewer to 4 more)
- Hospitalization for any cause: 35.6 fewer (95% CI 59.4 fewer to 11.8 more)
- Bradycardia: 20.1 more (95% CI 15.6 more to 24.6 more)
- Phosphenes (a visual adverse effect): 13.56 more (95% CI 9.90 more to 17.28 more)
- Atrial fibrillation: 7.26 more (95% CI 1.26 more to 13.2 more)

These estimates would result in a net effect estimate of a decrease in 9 (95% confidence interval decrease in 49 to increase in 32) hospitalization-equivalent events per 1000 patients.

This would be possible net benefit, and results in a low certainty of net benefit upon consideration across the range of relative importance that patients may have for the various effects.

Completing the Evidence-to-Decision Framework

A low certainty of net benefit supports a weak recommendation for ivabradine in patients meeting the selected criteria used in the trial suggesting benefit. Three of four current guidelines provide a weak recommendation for ivabradine in this setting (37-39) while one makes a strong recommendation (40).

Example 4. Second autologous stem cell transplant (ASCT) for patients with relapsed myeloma and response duration more than 2 years after first ASCT

A National Institute for Health and Clinical Excellence (NICE) guideline includes GRADE profiles for a second ASCT in relapsed myeloma including (41):

- Median overall survival from relapse – low quality evidence – absolute effect 2.1 years longer (95% CI not reported)
- Median time to progression – moderate quality evidence – absolute effect 13 months longer (95% CI not reported)
- No evidence identified for treatment-related morbidity and mortality, health-related quality of life, and adverse effects.

Step 1. Determine the outcomes to be combined.

The guideline panel considered overall survival and progression-free survival to be the most impactful outcomes for consideration. Because overall survival includes progression-free survival, progression events (time to progression) and death events (time to death) can be counted as non-overlapping outcomes.

Step 2. Determine the quantified relative importance for each outcome.

Time to death (overall survival) will be considered the reference unit. We will start with the assumption that patients would consider time to progression 0.2 times the importance of time to death.

Step 3. Combine the importance-adjusted effect estimates.

The importance-adjusted estimate for time to progression is median 2.6 months (0.2 x 13 months) and the estimate for time to death is median 2.1 years (or 25.2 months).

The point estimate for the net effect is the equivalent of 27.8 months of increased survival. Confidence intervals were not reported.

Step 4. Classify the precision of the net effect estimate.

This appears to start with a pattern of net benefit. The statistical significance was not expressed but assuming the results were statistically significant the confidence intervals would be completely within estimates of benefit because no evidence was provided to suggest harm.

Step 5. Consider the certainty of effect estimates for outcomes that are critical to the likelihood of net benefit.

Overall survival is the critical outcome here and was reported as low certainty of evidence based on a single retrospective comparative study (and related consistent data in noncomparative studies). Even so the guideline panel could potentially consider this to represent a moderate certainty of a survival

benefit and a low certainty for a specific magnitude of effect. This could lead to a moderate certainty of net benefit.

Step 6. Consider the range of relative importance for outcomes. Determine if the net effect estimate across the range of relative importance changes the certainty of net benefit rating.

In a model with no harms being considered as important outcomes, the range of relative importance is really an opportunity to consider how potential harms may affect the balance of benefits and harms. The guideline panel rationalized that harms would be similar to what patients experienced with their first ASCT and patients would thus have individual experience representing their individual harms estimates when considering the balance of harms and benefits.

Completing the Evidence-to-Decision Framework

In the context of harms mainly being considered burdens the patient would individually consider, the potential for increases in overall survival is considered a moderate certainty of net benefit. To reflect the importance of the patient weighing a personalized relative importance the guideline panel made a strong recommendation to offer the therapy (for the potential for net benefit) rather than recommend that the therapy should be administered.

Example 5. Avoiding 100% oxygen saturation in intensive care unit

A randomized trial with 480 adults admitted to the intensive care unit (ICU) found mortality in the ICU of 11.6% with target arterial oxyhemoglobin saturation (SpO2) 94%-98% versus 20.2% with target SpO2 97%-100% (absolute risk reduction 8.6%, 95% CI 1.7% to 15%) (42). This evidence may be considered to have moderate certainty due to early trial termination without use of a formal stopping rule.

Step 1. Determine the outcomes to be combined.

Although other outcomes were reported, there was no evidence of benefits for high-SpO2, so the mortality outcome can be considered the primary outcome for a net effect estimate.

Step 2. Determine the quantified relative importance for each outcome.

Step 3. Combine the importance-adjusted effect estimates.

These steps are irrelevant in this case and the net effect estimate is the estimate for ICU mortality, which can be considered inversely for the action of targeting an SpO2 97%-100% (absolute risk increase 8.6%, 95% confidence interval 1.7% to 15%).

Step 4. Classify the precision of the net effect estimate.

There is a net harm based on the confidence intervals of the effect estimate, consistent with a high certainty of net harm.

Step 5. Consider the certainty of effect estimates for outcomes that are critical to the likelihood of net benefit.

As the underlying evidence is a single trial with early unplanned termination, the moderate certainty of evidence may reduce the certainty of net harm to moderate.

Step 6. Consider the range of relative importance for outcomes. Determine if the net effect estimate across the range of relative importance changes the certainty of net benefit rating.

This is irrelevant following the decision to focus on a single outcome

Completing the Evidence-to-Decision Framework

A moderate certainty of net harm is sufficient to support a strong recommendation against an intervention with no apparent benefit.

Summary of Examples.

Five examples are presented to show how the model for defining and reporting the certainty of net benefit or certainty of net harm can provide more clear and explicit representations of evidence-based assessments and judgments supporting a recommendation spanning a broad continuum of complex situations. See Appendix Table 7.

Example 1 (Longer DAPT after drug-eluting stents) shows a net effect estimate suggesting a low certainty of net harm. Adjustment for certainty of evidence and the range of relative importance across outcomes is unnecessary as the certainty is already low. No other factors change the approach to the recommendation and we support a weak recommendation against. This may be clearer than the variations across current guidelines ranging from a weak recommendation for to a strong recommendation against.

Example 2 (Sacubitril-valsartan for symptomatic heart failure on standard therapy) is an example in which a net effect estimate shows net benefit based on a single trial using reasonable assumptions of relative importance of outcomes. The moderate certainty of evidence and the influence of reasonable extremes of relative importance assignments each led to ratings of a moderate certainty of net benefit. A moderate certainty of net benefit and high cost may support a weak recommendation for sacubitril-valsartan although many current guidelines provide a strong recommendation.

Example 3 (Ivabradine for symptomatic heart failure) shows a treatment with relatively smaller effects on benefits and harms with a closer balance between benefits and harms and moderate certainty of effect estimates. The resulting low certainty of net benefit supports a weak recommendation for ivabradine, and most current guidelines provide a weak recommendation for it.

Example 4 (Second ASCT for patients with relapsed myeloma and response duration more than 2 years after first ASCT) starts with low to moderate certainty in effect estimates for benefits and no direct comparative evidence to quantify harms. This leads to a higher certainty of net benefit, though still a moderate certainty of net benefit given the limited certainty in effect estimates. Despite not reaching a high certainty of net benefit the guideline panel reasoned that the harms for a second ASCT would be patient-specific (and patient-recognized based on the first ASCT) so provided a strong recommendation to offer the therapy and allow the patient to individually weigh the estimated benefits against their individualized harms. This is consistent with the GRADE approach which would provide a weak recommendation and encourage shared decision making. The guideline panel did not provide a strong recommendation for the intervention without shared decision making.

Example 5 (Avoiding 100% oxygen saturation in intensive care unit) is an example of an intervention with no apparent benefit and moderate certainty of net harm. The quantitative effect estimates support a high certainty of net harm but the risk of bias (qualitative certainty) reduced the overall assessment to a moderate certainty of net harm. Even so, without any apparent benefit, a strong recommendation against would be justified.

Appendix Table 7. Certainty of net benefit and strength of recommendations in examples

Example	Certainty of Evidence for Critical Outcomes*	Certainty of Net Benefit	Strength of Recommendation
Longer dual-antiplatelet therapy (DAPT) after drug-eluting stents	Moderate to high certainty of evidence	Low certainty of net harm	Weak recommendation against
Sacubitril-valsartan for symptomatic heart failure on standard therapy	Moderate certainty of evidence	Moderate certainty of net benefit	Weak recommendation for
Ivabradine for symptomatic heart failure	Moderate certainty of evidence	Low certainty of net benefit	Weak recommendation for
Second autologous stem cell transplant (ASCT) for patients with relapsed myeloma and response duration more than 2 years after first ASCT	Low to moderate certainty of evidence	Moderate certainty of net benefit	Weak recommendation for (or strong recommendation for offering with shared decision making)
Avoiding 100% oxygen saturation in intensive care unit	Moderate certainty of evidence	Moderate certainty of net harm	Strong recommendation against

* Certainty of evidence ratings here do not rate down for imprecision.

References

References 1-21 are the same as those included in the index article.

1. URL: www.gradeworkinggroup.org. Accessed November 27, 2016.
2. Atkins D, Best D, Briss PA, et al; GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ 2004;328:1490. doi: 10.1136/bmj.328.7454.1490.
3. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;64:401-6. doi: 10.1016/j.jclinepi.2010.07.015.
4. Hultcrantz M, Rind D, Akl E, et al; GRADE Working Group. The GRADE Working Group clarifies the construction of certainty of evidence. J Clin Epidemiol 2017;87:4-13. doi: 10.1016/j.jclinepi.2017.05.006.
5. Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol 2013;66:726-35. doi: 10.1016/j.jclinepi.2013.02.003.
6. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.3.2 Confidence in best estimates of magnitude of effects (quality of evidence). URL: gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.s1fsemn2yjpm. Accessed November 27, 2016.
7. Guyatt GH, Oxman AD, Kunz R, et al; GRADE Working Group. Going from evidence to recommendations. BMJ 2008;336:1049-51. doi: 10.1136/bmj.39493.646875.AE.
8. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.3.3 Confidence in values and preferences. URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.i5hfweocv3qs. Accessed November 27, 2016.
9. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.2.1 Balance of desirable and undesirable consequences. URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.feb5g22tf6nv. Accessed November 27, 2016.
10. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.3.4 Resource use (cost). URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.y0ey8465h47v. Accessed November 27, 2016.
11. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.5 The Evidence-to-Decision framework. URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.xr5ac2p2khuq. Accessed November 27, 2016.
12. Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D, U.S. Preventive Services Task Force. Update on the methods of the U.S. Preventive Services Task Force: estimating certainty and

magnitude of net benefit. Ann Intern Med 2007;147:871-5. doi: 10.7326/0003-4819-147-12-200712180-00007.

13. Krist AH, Wolff TA, Jonas DE, et al. Update on the methods of the U.S. Preventive Services Task Force: methods for understanding certainty and net benefit when making recommendations. Am J Prev Med 2018;54:S11-S18. doi: 10.1016/j.amepre.2017.09.011.
14. Hazlewood GS, Bombardier C, Tomlinson G, Marshall D. A Bayesian model that jointly considers comparative effectiveness research and patients' preferences may help inform GRADE recommendations: an application to rheumatoid arthritis treatment recommendations. J Clin Epidemiol 2018;93:56-65. doi: 10.1016/j.jclinepi.2017.10.003.
15. Fain KM, Yu T, Li T, et al. Evidence selection for a prescription drug's benefit-harm assessment: challenges and recommendations. J Clin Epidemiol 2016;74:151-7.
16. MacLean S, Mulla S, Akl EA, et al. Patient Values and Preferences in Decision Making for Antithrombotic Therapy: A Systematic Review: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 2012;141:e1S-e23S.
17. Richards T, Schroter S, Price A, Godlee F. Better together: patient partnership in medical journals. BMJ 2018;362:k3798.
18. Efron B, Tibshirani RJ. An Introduction to the Bootstrap, 1st Edition Chapman and Hall/CRC. New York, NY 1993
19. Spiegelhalter DJ. Myles JP, Jones DR, Abrams KR. An introduction to Bayesian methods in health technology assessment. BMJ 1999;319: 508–512.
20. Zhang Y, Alonso Coello P, et al. GRADE Guidelines: 19. Assessing the certainty of evidence in the importance of outcomes or values and preferences -- Risk of bias and indirectness. J Clin Epidemiol 2018.
21. Smith CM. Origin and uses of primum non nocere—above all, do no harm! J Clin Pharmacol 2005;45:371-7. doi: 10.1177/0091270004273680.
22. Schunemann HJ. Guidelines 2.0: do no net harm-the future of practice guideline development in asthma and other diseases. Current allergy and asthma reports 2011;11:261-8.
23. Spencer FA, Prasad M, Vandvik PO, et al. Longer versus shorter duration dual-antiplatelet therapy after drug-eluting stent placement: a systematic review and meta-analysis. Ann Intern Med 2015;163:118-26. doi: 10.7326/M15-0083.
24. Vandvik PO, Lincoff AM, Gore JM, et al. Primary and secondary prevention of cardiovascular disease: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 2012;141:e637S-e666S. doi: 10.1378/chest.11-2306.
25. Levine GN, Bates ER, Blankenship JC, et al; American College of Cardiology Foundation; American Heart Association Task Force on Practice Guidelines; Society for Cardiovascular Angiography and Interventions. 2011 ACCF/AHA/SCAI Guideline for Percutaneous Coronary Intervention. A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines and the Society for Cardiovascular Angiography and Interventions. J Am Coll Cardiol 2011;58:e44-122. doi: 10.1016/j.jacc.2011.08.007.
26. DynaMed Plus [Internet]. Ipswich (MA): EBSCO Information Services. 1995-2016. Record No. 114099, Heart failure with reduced ejection fraction; [updated 2016 Oct 07, cited 2016 Nov 27]; [about 101 screens]. Available from http://www.dynamed.com/topics/dmp~AN~T114099/Heart-failure-with-reduced-ejection-fraction#Sacubitril-valsartan--ARNI-.
27. Database of Evidence Profiles [Internet]. Alper B, Colpitts C, Lang E, et al. Should sacubitril-valsartan vs. ACE inhibitor or ARB be used for symptomatic patients with heart failure and

reduced ejection fraction? [updated 2015 Nov 17; cited 2016 Nov 27]. Available from http://dbep.gradepro.org/profile/1FEB6C9D-1F7A-EAE1-AEC5-564B423DF770.

28. McMurray JJ, Packer M, Desai AS, et al; PARADIGM-HF Investigators and Committees. Angiotensin-neprilysin inhibition versus enalapril in heart failure. N Engl J Med 2014;371:993-1004. doi: 10.1056/NEJMoa1409077.
29. WRITING COMMITTEE MEMBERS; Yancy CW, Jessup M, Bozkurt B, et al. 2016 ACC/AHA/HFSA Focused Update on New Pharmacologic Therapy for Heart Failure: An Update of the 2013 ACCF/AHA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. Circulation 2016;134:e282-93. doi: 10.1161/CIR.0000000000000435.
30. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al; Authors/Task Force Members; Document Reviewers. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur J Heart Fail 2016;18:891-975. doi: 10.1002/ejhf.592.
31. Scottish Intercollegiate Guidelines Network (SIGN). National clinical guideline for management of chronic heart failure. SIGN 2016 Mar. SIGN publication no. 147.
32. National Institute for Health and Care Excellence (NICE). Sacubitril valsartan for treating symptomatic chronic heart failure with reduced ejection fraction. Technology appraisal guidance. Published: 27 April 2016. TA388.
33. Fox K, Ford I, Steg PG, Tendera M, Ferrari R; BEAUTIFUL Investigators. Ivabradine for patients with stable coronary artery disease and left-ventricular systolic dysfunction (BEAUTIFUL): a randomised, double-blind, placebo-controlled trial. Lancet 2008;372:807-16. doi: 10.1016/S0140-6736(08)61170-8.
34. Swedberg K, Komajda M, Bohm M, et al; SHIFT Investigators. Ivabradine and outcomes in chronic heart failure (SHIFT): a randomised placebo-controlled study. Lancet 2010;376:875-85. doi: 10.1016/S0140-6736(10)61198-1.
35. ClinicalTrials.gov. Effects of Ivabradine on Cardiovascular Events in Patients With Moderate to Severe Chronic Heart Failure and Left Ventricular Systolic Dysfunction. A Three-year International Multicentre Study (SHIFT). NCT02441218. Last Updated May 9, 2016.
36. ClinicalTrials.gov. The BEAUTIFUL Study: Effects of Ivabradine in Patients With Stable Coronary Artery Disease and Left Ventricular Systolic Dysfunction. NCT00143507. Last Updated June 22, 2015.
37. Writing Committee Members, Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr, et al. 2016 ACC/AHA/HFSA Focused Update on New Pharmacological Therapy for Heart Failure: An Update of the 2013 ACCF/AHA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. Circulation 2016;134:e282-93. doi: 10.1161/CIR. 0000000000000435.
38. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al; Authors/Task Force Members. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2016;37:2129-200. doi: 10.1093/eurheartj/ehw128.

39. National Institute for Health and Care Excellence (NICE). Ivabradine for treating chronic heart failure. Technology appraisal guidance. Published: 28 November 2012. TA267.
40. Scottish Intercollegiate Guidelines Network (SIGN). Management of chronic heart failure: A national clinical guideline. Published: March 2016. SIGN publication no. 147. http://www.sign.ac.uk/assets/sign147.pdf
41. National Collaborating Centre for Cancer. Myeloma: diagnosis and management. NICE Guideline 35, Full guideline, February 2016. https://www.nice.org.uk/guidance/ng35/evidence/full-guideline-2306487277
42. Giardis M, Busani S, Damiani E, et al. Effect of conservative vs conventional oxygen therapy on mortality among patients in an intensive care unit: The Oxygen-ICU Randomized Clinical Trial. JAMA 2016;316:1583-1589. doi: 10.1001/jama.2016.11993.
43. Krumholz HM, Lin Z, Keenan PS, et al. Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. JAMA 2013;309:587-93.

BMJ Open

Defining Certainty of Net Benefit: a GRADE concept paper

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2018-027445.R2
Article Type:	Communication article
Date Submitted by the Author:	18-Apr-2019
Complete List of Authors:	Alper, Brian; EBSCO Health, DynaMed Plus; University of Missouri Columbia School of Medicine, Family and Community Medicine Oettgen, Peter; EBSCO Health, DynaMed Kunnamo, Ilkka; Duodecim Medical Publications Ltd.; The Finnish Medical Society Iorio, Alfonso; McMaster University, Health Research Methods, Evidence, and Impact Ansari, Mohammed; School of Epidemiology, Public Health and Preventive Medicine, Faculty of Medicine, University of Ottawa Murad, M. Hassan; Mayo Clinic, Evidence-Based Practice Center Meerpohl, Joerg; Institute of Medical Biometry & Medical Informatics, University Medical Centre Freiburg, German Cochrane Centre; University Medical Centre, Paediatric Haematology & Oncology, Centre for Paediatrics & Adolescent Medicine Qaseem, Amir; American College of Physicians, Department of Clinical Policy Hultcrantz, Monica; Karolinska Institute; Statens beredning for medicinsk utvardering Schünemann, Holger; McMaster University, Departments of Health Research Methods, Evidence, and Impact and of Medicine, McMaster University Health Sciences Centre, Hamilton, Canada Guyatt, Gordon; Mcmaster University, Health Research Methods, Evidence, and Impact GRADE Working Group, The; GRADE Working Group
Primary Subject Heading:	Evidence based practice
Secondary Subject Heading:	Communication, Epidemiology, Health policy
Keywords:	evidence-based medicine, decision analysis, evidence synthesis, clinical decision making, guideline development

Defining Certainty of Net Benefit: a GRADE concept paper

GRADE defines Certainty of Net Benefit

Brian S Alper, Peter Oettgen, Ilkka Kunnamo, Alfonso Iorio, Mohammed T Ansari, M Hassan Murad, Joerg J Meerpohl, Amir Qaseem, Monica Hultcrantz, Holger Schünemann, Gordon Guyatt, the GRADE Working Group

Brian S Alper, Vice President of Innovations and Evidence-Based Medicine Development, EBSCO Health, EBSCO Information Services, Inc., Ipswich, Massachusetts, United States; Clinical Research Assistant Professor, Department of Family and Community Medicine, University of Missouri-Columbia School of Medicine

Peter Oettgen, Editor in Chief, DynaMed Plus, EBSCO Health, EBSCO Information Services, Inc., Ipswich, Massachusetts, United States

Ilkka Kunnamo, Research Director, Duodecim Medical Publications Ltd.

Alfonso Iorio, Professor, Health Research Methods, Evidence and Impact, McMaster University

Mohammed T Ansari, Adjunct Professor, School of Epidemiology and Public Health, Faculty of Medicine, University of Ottawa

M Hassan Murad, Evidence-based Practice Center, Mayo Clinic

Joerg J Meerpohl, Institute for Evidence in Medicine (for Cochrane Germany Foundation); Medical Center – University of Freiburg

Amir Qaseem, Vice President for Clinical Policy and the Center for Evidence Reviews, American College of Physicians

Monica Hultcrantz, Swedish Agency for Health Technology Assessment and Assessment of Social Services; Karolinska Institutet

Holger J Schünemann, Chair, Health Research Methods, Evidence and Impact, McMaster University; Director, Cochrane Canada

Gordon Guyatt, Professor, Health Research Methods, Evidence and Impact, McMaster University

The GRADE Working Group

Correspondence to Brian S. Alper, balper@ebsco.com

Word count: 3581

Abstract

Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology is used to assess and report certainty of evidence and strength of recommendations. This GRADE concept article is not GRADE guidance but introduces certainty of net benefit, defined as the certainty that the balance between desirable and undesirable health effects is favorable. Determining certainty of net benefit requires considering certainty of effect estimates, the expected importance of outcomes and variability in importance, and the interaction of these concepts. Certainty of net harm is the certainty that the net effect is unfavorable. Guideline panels using or testing this approach might limit strong recommendations to actions with a high certainty of net benefit or against actions with a moderate or high certainty of net harm. Recommendations may differ in direction or strength from that suggested by the certainty of net benefit or harm when influenced by cost, equity, acceptability, or feasibility.

Introduction

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group has designed a transparent approach to rating certainty of evidence and grading strength of recommendations (1,2). More than one hundred groups creating systematic reviews, clinical practice and public health guidelines, and health technology assessments have adopted GRADE (1,2). GRADE uses the terms "certainty of evidence" interchangeably with "confidence in estimate" and "quality of evidence". Authors using GRADE make separate ratings of certainty for each patient-important outcome and, in the context of a recommendation about an intervention, provide an overall rating based on the lowest certainty of the critical outcomes.

In the context of making recommendations, GRADE specifies that ratings reflect the certainty that the estimates of an effect are adequate to support a particular decision or recommendation (3). Recently, the GRADE Working Group clarified the conceptual basis of certainty ratings, noting that, in both contexts of systematic reviews and guidelines, they represent the certainty that a true effect lies on one side of a specified threshold, or within a specified range (4).

Depending on the thresholds or ranges chosen, it is possible to have high certainty in the evidence for a set of outcomes related to a particular decision, yet uncertainty whether the evidence is adequate to support that decision; this will occur when desirable and undesirable consequences are closely balanced, such as cancer treatments with high certainty in prolonging survival and high certainty in serious toxicity (5,6). It is also possible to have low certainty in evidence for a specific outcome yet make a strong recommendation (high certainty to support a decision). The GRADE Working Group has specified five paradigmatic situations in which such discordant recommendations may be appropriate (5,6). One of these situations is when only low quality evidence exists for a promising intervention in a life-threatening context (e.g. using fresh frozen plasma or vitamin K in a patient receiving warfarin with elevated INR and an intracranial bleed).

The recent GRADE Working Group guidance states that systematic review authors and guideline panelists will ideally specify the threshold or ranges they are using when rating the certainty in evidence (3). The guidance offered non-contextualized (no implicit value judgments) and partially contextualized (some implicit value judgments regarding magnitude of effects) approaches for systematic review authors. The guidance further suggested a fully contextualized approach for clinical practice guidelines in which a guideline panel determines thresholds considering all critical outcomes and their relative importance.

Guideline panels using fully contextualized approaches have faced challenges of balancing feasibility and simplicity with comprehensive simultaneous consideration of all important outcomes. This current GRADE concept article introduces an approach for guideline panels to more directly and explicitly rate their certainty of the balance of benefits and harms. This GRADE concept article (a new form of communication from the GRADE Working Group) is presented to stimulate discussion and does not constitute GRADE guidance.

Expressing Certainty Across the Evidence-to-Decision Framework

GRADE Evidence-to-Decision frameworks explicitly identify the following considerations in determining the direction and strength of recommendations:

- Certainty of evidence (regarding effect estimates for health effects) (2,6,7)
- Relative importance of outcomes (also called values and preferences) (2,7,8)
- Balance of benefits and harms (2,7,9)
- Resource use (cost) (2,7,10)
- Cost-benefit ratio (Are incremental health benefits worth the costs?) (2,7,11)
- Equity (7,11)
- Acceptability (11), and
- Feasibility (11).

Health-related harms include pain or disability but also burdens that lower quality of life. For example, the burden of receiving an intervention that requires being immobile for long periods of time could be considered as a health-related harm. In this article, when we use the phrase "balance of benefits and harms" we refer to the "balance of benefits versus harms and burdens". Other burdens that may be considered more societal in nature may be considered through other criteria in the framework (cost, acceptability, feasibility) depending on the perspective taken such as that of the health care system, the population or the individual. Here we will use the term "harms" to refer to "health-related harms and burdens".

Ideally, guideline panels consider all the factors listed above when determining the direction and strength of a recommendation. The process may proceed in progressive steps that consider first benefits and harms to generate certainty in net benefit; then costs to generate certainty in a cost-benefit ratio; then equity, acceptability, and equity to address certainty in a recommendation if relevant (Figure 1).

Although it makes decisions more transparent, reporting a guideline panel's certainty for each of these concepts may be overwhelming for guideline users seeking simple explanations of the rationale and certainty for recommendations. Among the concepts for which certainty can be expressed formally, the certainty in balance of benefits and harms (net effect) may be most relevant to patients and clinicians (often the primary target users for guidelines). Additional criteria that may influence a recommendation (cost, cost-benefit ratio, equity, acceptability, feasibility) are more likely to vary across social groups and contexts, and population-based ratings may be of less interest to patients and clinicians working together to make individual health care decisions.

Consistent with the recent clarification of "certainty of evidence" - the certainty that a true effect lies within a specified range or on one side of a specified threshold (3) - one can express the certainty of the net effect (or balance of benefits and harms) in terms of a range or in relation to a threshold. The situation when benefits and harms are perfectly balanced (net benefit or harm = 0) represents a natural threshold for certainty of the net effect. Using this threshold, the certainty of net benefit is the certainty

that the overall or net effect lies on the side of benefit. The certainty of net harm is the certainty that the net effect lies on the side of harm.

Expressing the certainty of net benefit for guideline users provides the most direct summary representation of the extent of our confidence that the estimates of effects are adequate to support a particular decision or recommendation. The United States Preventive Services Task Force (USPSTF) has used the term certainty of net benefit in a manner consistent with this conceptual framework (12, 13).

Model for Creating the Net Effect Estimate and Rating Its Certainty

Determining the certainty in the balance of benefits and harms involves generating a net effect estimate (a way of specifying the balance of benefits and harms) and then rating the certainty regarding that net effect in relation to the threshold of net benefit = 0 (Figure 2).

Decision analysis provides a statistical method for generating the net effect estimate. Decision modeling has evolved over the years and sophisticated models include multiple outcomes, the varying times at which each outcome can occur, the relative importance placed in each outcome (often using utilities or quality adjusted life years), and future decisions and resulting outcomes. Guideline panels sometimes use decision analysis to evaluate a chain of possible consequences and decisions to inform their recommendations: the UK National Institute for Health and Care Excellence (NICE) relies heavily on such models. Decision analysis often involves modelling cost-effectiveness or an assessment of net effect across a range of possible scenarios. Determining the certainty of evidence emerging from such models is itself a complex matter: A GRADE project group is currently addressing the issue.

For many decisions for which guideline developers, clinicians or patients desire recommendations, however, one need not consider a chain of subsequent decisions. Many guideline recommendations are binary and based on the evidence limited to that decision. In such cases, one can perform a much simpler decision analysis without requiring participation of a skilled modeler. Simple models can generate confidence intervals for a net effect estimate (a composite of individual effect estimates) given the following assumptions (described further in the Appendix):

1. Effect estimates represent data conforming to normal distributions
2. Effect estimates to be combined are independent and not correlated with each other
3. Effect estimates to be combined can be multiplied by a conversion factor to use a consistent unit of measure

Given that the second assumption is often unlikely to hold, the analyst can perform sensitivity analysis of the net effect estimate to determine robustness to changes in the individual effect estimates, the assumptions of correlation between effect estimates, and the conversion factors. A sensitivity analysis defining the likelihood of the net effect estimate remaining favorable across the range of assumptions determines the certainty of net benefit.

Generation of the Net Effect Estimate

Here we describe the methods for generating the net effect estimate as presented in Figure 2. Algorithm-supported calculators can facilitate combining the importance-adjusted effect estimates (the third step in Figure 2) and classifying the precision (the fourth step). The Appendix provides examples and a link to a free online calculator.

Step 1: Determine the outcomes to be combined

We assume reviewers have already identified the important outcomes for their systematic review of the available evidence; methods for this outcome selection have been reported (14). We present here considerations for selecting from those outcomes the outcomes to be combined for a net effect estimate.

Including both a composite outcome and one or more components of that outcome is problematic. For example, it would be inappropriate to include all-cause mortality and cardiovascular mortality in the same model. One may choose to use only the composite outcome (e.g. all-cause mortality) or to use only the component outcomes (e.g. cardiovascular mortality, cancer mortality, and mortality from causes other than cancer or cardiovascular disease).

If effect estimates are not available in absolute terms (or if effect estimates are being extrapolated to a population with different baseline risks than that used for the absolute effect estimates) then absolute effect estimates may be derived using a combination of relative effect estimates and baseline risk estimates.

Step 2: Determine the quantified relative importance for each outcome

Quantitative estimates of relative importance for each outcome will serve as a conversion factor to use a consistent unit of measure for the net effect estimate. These estimates need to be meaningful as a multiplier or represent a quantitative measure of importance relative to a reference standard. Guideline panels that use a qualitative 9-point rating of importance of outcomes(14) to determine which outcomes to include in systematic reviews or summary of findings tables may find these ratings do not easily translate to quantitative estimates for this purpose.

A simple approach is to select one outcome as a reference outcome and define a relative importance adjustment (i.e., a multiplier) for each other outcome as a modifier to apply to effect estimates. In making individual patient-specific decisions, one could enter the quantitative estimates of relative importance for the individual patient and derive an individualized estimate of net effect. With further development this approach could inform shared decision-making for individual patients.

For groups of patients, one could consider quantitative estimates of relative importance as ranges. In making population-specific recommendations, one could use a range of relative importance estimates considered reasonable to capture most members of the population and check for robustness of estimates of net effect across the range of relative importance. One would then lower the rating of certainty of net benefit if the estimate of net effect crosses to net harm within the range of relative

importance. The later discussion of sensitivity analysis for the net effect estimate (step 6) will address the concepts of ranges and certainties of relative importance.

Methods to determine quantitative estimates of relative importance from a patient perspective include discrete-choice experiments (15), preference-eliciting surveys among patients (16), and systematic reviews of such surveys (17). Determination of relative importance could provide an opportunity for engaging patients as partners in research design, a developing expectation in medical publishing (18). When such evidence is unavailable for the outcomes associated with a recommendation, guideline panels can still explicitly make best guesses of the importance the target population will place on the relevant outcomes. Further discussion of the methods for determining relative importance is beyond the scope of this paper.

If the outcomes to be combined include both continuous measures and dichotomous measures, the assignment of relative importance becomes more complicated and would take additional methods to reach a shared unit of measure (such as conversion to quality-adjusted life-year estimates). Utilities reported for decision analyses may be convertible to relative importance of outcomes. However, utilities are often reported with a range from 0 (for death or worst outcome) to 1 (for optimal quality of life or best outcome), and relative importance of outcomes functioning as multipliers would not be meaningful if multiplied by 0. Relative importance of outcome estimates equal to 1 minus the utility could convert utilities to meaningful multipliers.

Step 3: Combine the importance-adjusted effect estimates

For each effect estimate, one can multiply the point estimate and confidence intervals (CI) by the relative importance for the outcome, and then present the importance-adjusted effect estimate in positive or negative terms to correspond to benefits or harms in the direction of effect.

Adding together the point estimates for each importance-adjusted effect estimate will provide the point estimate for the net effect. Statistical formulas allow calculation of the 95% CI for the net effect (see Appendix).

Rating the Certainty of Net Benefit

Step 4: Classify the precision of the net effect estimate

Precision becomes meaningful with contextual anchoring. Reporting results with a 3-centimeter range would be overly precise for planning travel by car and unacceptable imprecision for some types of surgery. To express the certainty in the balance of benefits and harms, we need to specify a threshold for a net benefit, then express the certainty that the net effect lies on one side of this threshold.

Guideline panels may specify the threshold of net effect; we suggest using the "zero effect" for simplicity. Guideline panels that formally evaluate cost-effectiveness already use a method to set a value threshold for the quantity of net benefit that is considered worth the cost to achieve it.

If the entire confidence interval does not cross zero, then the precision of the net effect estimate is sufficient to not rate down the certainty of net benefit for imprecision. One must still consider other factors affecting certainty that are more difficult to quantify (risk of bias, inconsistency, indirectness, and publication bias) and the plausible range of relative importance of outcomes before final determination of the certainty of net benefit (19).

If the confidence interval includes zero effect and thus the range of net effect estimates includes both net benefit and net harm, the guideline panel will rate down the certainty of net benefit. The greater the extent of overlap of the confidence interval with both benefit and harm, the lower the certainty in the net benefit. Table 1 and Figure 3 present initial suggestions for how these judgments may be made.

Table 1. Classification of precision of net effect estimate

Pattern of net effect estimate	Classification	Precision of net effect estimate is consistent with …
Entire confidence interval is beneficial	Net benefit	High certainty of net benefit
Point estimate is beneficial, lower bound of confidence interval is harmful, and point estimate has larger absolute value than lower bound of confidence interval	Likely net benefit	Moderate certainty of net benefit
Point estimate is beneficial, lower bound of confidence interval is harmful, and point estimate has smaller absolute value than lower bound of confidence interval	Possible net benefit	Low certainty of net benefit
Point estimate is close to zero, wide confidence interval*	Possibly no net benefit or harm	Very low certainty of net benefit or harm
Point estimate is close to zero, narrow confidence interval*	Net benefit or harm likely near zero	Moderate certainty of little net benefit or harm
Point estimate is harmful, upper bound of confidence interval is beneficial, and point estimate has smaller absolute value than upper bound of confidence interval	Possible net harm	Low certainty of net harm
Point estimate is harmful, upper bound of confidence interval is beneficial, and point estimate has larger absolute value than upper bound of confidence interval	Likely net harm	Moderate certainty of net harm
Entire confidence interval is harmful	Net harm	High certainty of net harm

* Differentiation of wide vs. narrow confidence intervals could be based on a threshold of minimally important differences.

The calculation for confidence intervals for the net effect estimate includes an assumption that effect estimates being combined are not correlated with each other. If effects are correlated, the accurate confidence intervals would be wider or less precise; if inversely correlated, the accurate confidence intervals would be narrower or more precise. If such accuracy is needed, one could add correlation coefficients to the calculation (see Appendix) or rely on more sophisticated statistical approaches such as bootstrapping (20) or a Bayesian approach to estimate the probability interval (21). The calculation is also based on an assumption that effects on outcomes are independent. For practical use, modest violations of the assumption are unlikely to distort results substantially and may be preferable to less explicit judgment of the balance of benefits and harms.

Step 5: Consider the certainty of effect estimates for outcomes that are critical to the likelihood of net benefit.

One approach to select the outcomes critical to the likelihood of net benefit is to identify the outcomes that could change the classification of the precision of the net effect estimate. Such outcomes are either:

- Outcomes for which removal of the outcome would change the classification of the precision of the net effect estimate.
- Outcomes for which addition of plausible increases to the effect estimate (for effect estimates with lower certainty) would change the classification.

Determining the lowest certainty of evidence among critical outcomes requires addressing risk of bias, inconsistency, indirectness, and publication bias for each critical outcome(4). Imprecision for an individual outcome is not an influencing factor here because it is already accounted for in the net effect estimate.

The lowest of the certainty ratings for critical outcomes and the certainty rating consistent with the precision of the net effect estimate represents the certainty of net benefit. This approach may work in most cases; raters still need, however, to consider the overall framework and determine if limited certainty in single outcomes are sufficient to rate down the overall certainty of net benefit. This is especially so if the upper or lower bounds of the CI for the net effect estimate approximates a zero effect. A 95% CI is used based on convention rather than a theoretical rationale.

Step 6: Consider the range of relative importance for outcomes. Perform a sensitivity analysis to determine the certainty of net benefit across this range.

To enhance feasibility of the approach, efforts to fully consider the range of relative importance for outcomes may be limited to ratings that would otherwise be classified as high certainty of net benefit. In situations in which further assessment is needed to confirm robustness of certainty across the range of relative importance, one can repeat the analyses across a reasonable range of relative importance of outcomes.

The purpose of the sensitivity analysis is to determine if the certainty of net benefit remains high across the range of relative importance estimates. There remains insufficient conceptual development to provide explicit guidance on how to precisely define the range of relative importance for outcomes to use for the sensitivity analysis.

The GRADE Working Group has developed guidance on rating the certainty of relative importance of outcomes (22). If a range of relative importance of outcomes is determined by empirical evidence and that range is considered to have low certainty, it would then be prudent to use a wider range of relative importance of outcomes in a sensitivity analysis.

It may be necessary during the process of the sensitivity analysis of outcome importance to re-evaluate which outcomes are critical to the likelihood of net benefit.

Relating Certainty of Net Benefit to Strength of Recommendation

The certainty of net benefit does not necessarily dictate the strength of recommendation. The evidence-to-decision framework also includes cost, cost-benefit ratio, equity, acceptability, and feasibility as considerations that may modify the strength of recommendation. Panels may choose to focus exclusively on net health effects and not include other elements (e.g. some panels choose not to consider costs, and do not formally consider acceptability, feasibility, and equity).

In situations in which there is a high certainty in effect estimates but uncertainty that the balance of benefits and harms is favorable across the range of patient values and preferences (a situation in which panels will make weak recommendations because fully informed patients are likely to make different decisions), a moderate or low certainty of net benefit provides a clear expression of the rationale for weak recommendations.

High certainty is not necessary, in all cases, for supporting a strong recommendation. *Primum non nocere* ("First, do no harm") is considered one of the principal precepts for ethical decision-making in medicine and pharmacology (23) though it is more properly considered *Primum non net nocere* (24). One can interpret this to consider a lower threshold for the certainty in net harm for a strong recommendation against an action than one uses for the certainty in net benefit for a strong recommendation for an action.

Implications

In this article, we introduce an approach for guideline developers to consider explicitly reporting the certainty of net benefit with recommendations, either in addition to or in place of reporting an overall quality of evidence associated with a recommendation. Either way, the approach requires consideration of certainty of evidence ratings for individual outcomes, typically presented in summary of findings tables.

This approach is applicable to decisions or recommendations with binary choices, such as treatment, prevention, diagnostic and screening interventions. This approach involves many judgments that are already made explicitly or implicitly when guideline panels make recommendations. Reporting the

judgments made when using this approach would allow readers to interpret their confidence in how the ratings were made and may reduce spurious confidence that could occur with quantitative reporting in the absence of qualitative factors.

A key driver for this approach is greater congruence with the intent behind the concept of "adequate evidence to support a recommendation" than what is currently conveyed by the "overall quality of evidence in estimates of effects". Strengths of this approach include the transparent, logical, quantitative expressions for both scholarly and clinical readers, and both guideline developers and guideline users.

Throughout this discussion we are considering the context of guideline recommendations which by nature relate to considerations for a population and not for a specific individual. Concepts of certainty of net benefit may eventually be extrapolated to "certainty of individual net benefit" with inclusion of individually determined relative importance of outcomes, but at this time no discussion or testing has been applied to relating these concepts to individual decision making.

The primary limitation of this approach is its lack of testing to inform its feasibility and acceptability, and how readers will interpret these concepts. This report is shared, before such testing, to increase scholarly discussion. This GRADE concept article does not therefore constitute GRADE guidance.

Acknowledgements: The authors are pleased to acknowledge Lehana Thabane, PhD for collaborating with the lead author to derive the statistical method; Paul E. Alexander, MSc, MHSc, Cynthia Boyd, MD, MPH, Reem Mustafa, MD and Irfan Dhalla MD, MSc for feedback to improve the readability and conceptual clarity of the manuscript; and Jordan Prince for construction of an online calculator.

Contributorship statement: At a GRADE Working Group (GWG) meeting GG and MH presented a complex method to rate the certainty of evidence for an outcome when fully contextualized with respect to other outcomes related to a decision or recommendation. HJS had introduced the concept of rating certainty in a range of effects based on all GRADE domains. BSA introduced the concept of certainty of net benefit to clarify and simplify methodology to report and assess the balance of benefits and harms in the context of fully contextualizing certainty of evidence across outcomes. The GWG formed a subcommittee for this conceptual development. BSA and Lehana Thabane developed the statistical model for a simple decision analysis and sensitivity analysis. BSA, IK, AI and Lehana Thabane provided examples to demonstrate the model (Appendix). The GWG had in-person meetings in three countries (with up to 100 people in attendance) in which the wider audience provided in-depth review, feedback and discussion to refine the concepts. BSA, PO, IK, AI, MTA, MHM, JJM, AQ, MH, HS and GG met frequently to iteratively refine the concepts, meet the authorship requirements, and approve the final version. BSA is the guarantor of the article.

Funding: No funding support.

Competing interests: All authors have completed the unified competing interest form at www.icmje.org/coi_disclosure.pdf (available on request from the corresponding author). There are no direct competing interests as the article provides conceptual development and no promotion or assessment of any product or service. All authors are members of the GRADE Working Group and conduct scholarly activity or professional services related to the concepts in this article. BSA and PO are employed by EBSCO Information Services and IK is employed by Duodecim Medical Publications Ltd.

Transparency: BSA affirms that that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as originally planned (and, if relevant, registered) have been explained.

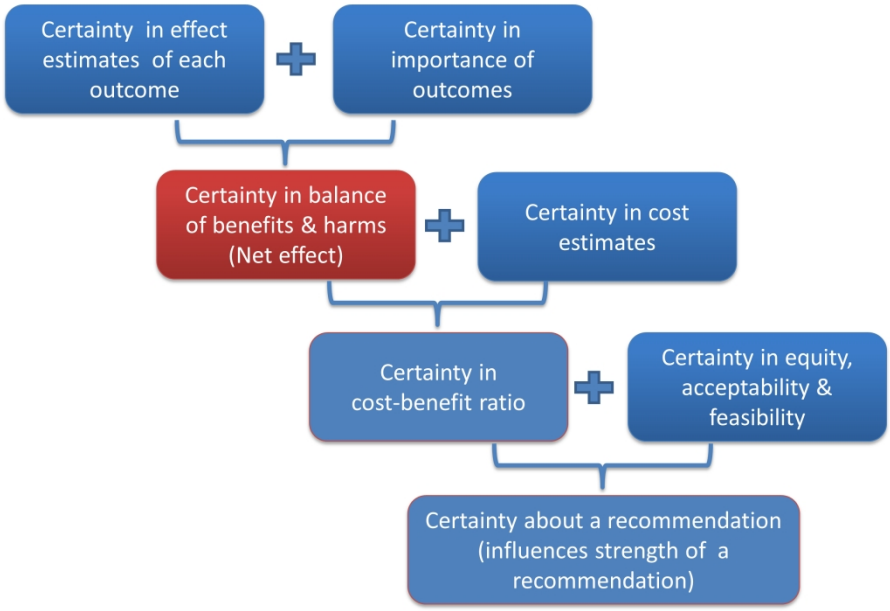
Data Sharing Statement: No additional data available

Patient and Public Involvement statement: The manuscript was shared with the Patient Editor for Research and Evaluation at The BMJ (Amy Price) who welcomed the development and suggested Patient Involvement was not directly appropriate for this report. Public Involvement was not openly pursued but the GRADE Working Group provided multiple opportunities for more than 100 multidisciplinary stakeholders to contribute.

References:

1. URL: www.gradeworkinggroup.org. Accessed November 27, 2016.
2. Atkins D, Best D, Briss PA, et al; GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ 2004;328:1490. doi: 10.1136/bmj.328.7454.1490.
3. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;64:401-6. doi: 10.1016/j.jclinepi.2010.07.015.
4. Hultcrantz M, Rind D, Akl E, et al; GRADE Working Group. The GRADE Working Group clarifies the construction of certainty of evidence. J Clin Epidemiol 2017;87:4-13. doi: 10.1016/j.jclinepi.2017.05.006.
5. Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol 2013;66:726-35. doi: 10.1016/j.jclinepi.2013.02.003.
6. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.3.2 Confidence in best estimates of magnitude of effects (quality of evidence). URL: gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.s1fsemn2yjpm. Accessed November 27, 2016.
7. Guyatt GH, Oxman AD, Kunz R, et al; GRADE Working Group. Going from evidence to recommendations. BMJ 2008;336(7652):1049-51. doi: 10.1136/bmj.39493.646875.AE.
8. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.3.3 Confidence in values and preferences. URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.i5hfweocv3qs. Accessed November 27, 2016.
9. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.2.1 Balance of desirable and undesirable consequences. URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.feb5g22tf6nv. Accessed November 27, 2016.
10. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.3.4 Resource use (cost). URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.y0ey8465h47v. Accessed November 27, 2016.
11. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.5 The Evidence-to-Decision framework. URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.xr5ac2p2khuq. Accessed November 27, 2016.
12. Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D, U.S. Preventive Services Task Force. Update on the methods of the U.S. Preventive Services Task Force: estimating certainty and magnitude of net benefit. Ann Intern Med 2007;147:871-5. doi: 10.7326/0003-4819-147-12-200712180-00007.

13. Krist AH, Wolff TA, Jonas DE, et al. Update on the methods of the U.S. Preventive Services Task Force: methods for understanding certainty and net benefit when making recommendations. Am J Prev Med 2018;54:S11-S18. doi: 10.1016/j.amepre.2017.09.011.
14. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 3 Selecting and rating the importance of outcomes. URL: https://gdt.gradepro.org/app/handbook/handbook.html#h.1i2bwkm8zpjo. Accessed April 6, 2019.
15. Hazlewood GS, Bombardier C, Tomlinson G, Marshall D. A Bayesian model that jointly considers comparative effectiveness research and patients' preferences may help inform GRADE recommendations: an application to rheumatoid arthritis treatment recommendations. J Clin Epidemiol 2018;93:56-65. doi: 10.1016/j.jclinepi.2017.10.003.
16. Fain KM, Yu T, Li T, Boyd CM, Singh S, Puhan MA. Evidence selection for a prescription drug's benefit-harm assessment: challenges and recommendations. J Clin Epidemiol 2016;74:151-7.
17. MacLean S, Mulla S, Akl EA, et al. Patient Values and Preferences in Decision Making for Antithrombotic Therapy: A Systematic Review: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 2012;141:e1S-e23S.
18. Richards T, Schroter S, Price A, Godlee F. Better together: patient partnership in medical journals. BMJ 2018;362:k3798.
19. Schünemann HJ. Interpreting GRADES's levels of certainty or quality of the evidence: GRADE for statisticians, considering review information size or less emphasis on imprecision? J Clin Epidemiol 2016;75:6-15.
20. Efron B, Tibshirani RJ. An Introduction to the Bootstrap, 1st Edition Chapman and Hall/CRC. New York, NY 1993.
21. Spiegelhalter DJ. Myles JP, Jones DR, Abrams KR. An introduction to Bayesian methods in health technology assessment. BMJ 1999;319:508–512.
22. Zhang Y, Alonso Coello P, Guyatt G, Yepes-Nunez JJ, Akl EA, Hazlewood G, et al. GRADE Guidelines: 19. Assessing the certainty of evidence in the importance of outcomes or values and preferences -- Risk of bias and indirectness. J Clin Epidemiol 2018.
23. Smith CM. Origin and uses of primum non nocere—above all, do no harm! J Clin Pharmacol 2005;45:371-7. doi: 10.1177/0091270004273680
24. Schünemann HJ. Guidelines 2.0: do no net harm-the future of practice guideline development in asthma and other diseases. Current allergy and asthma reports 2011;11:261-8.



* This figure is not intended to convey a hierarchy, order or requirement of factors for consideration. The red component represents the key concept of this discussion.

Figure 1. Certainty across the evidence-to-decision framework*

254x190mm (300 x 300 DPI)

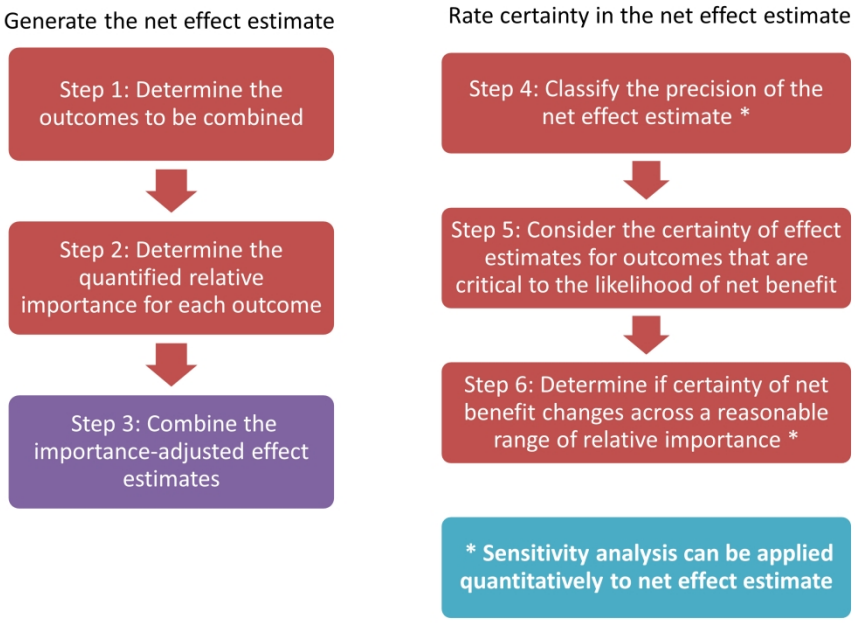


Figure 2. A stepwise approach to determining the certainty of the net effect estimate

254x190mm (300 x 300 DPI)

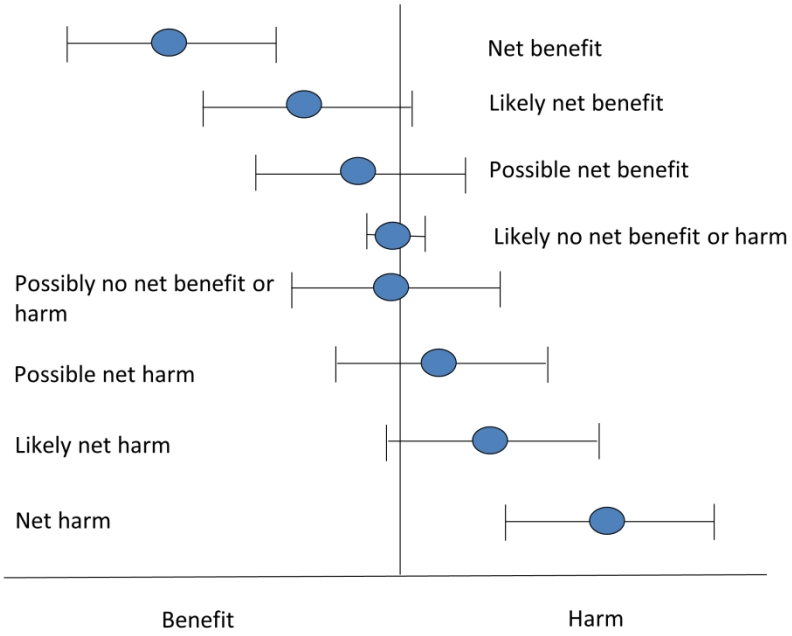


Figure 3. Classification of precision of net effect estimate

254x190mm (300 x 300 DPI)

Supplemental Material (Appendix)

Calculating the Net Effect Estimate

The statistical method is fully described here and a free online calculator is available at ebscohealth.com/innovations. In brief, while we add the point estimates for each effect estimate to determine the point estimate for the net effect estimate, the calculation of the 95% confidence interval requires a couple of formulas.

Suppose we have an effect estimate X for one outcome and Y for another outcome and we want to determine a combined or net effect estimate Z. The model to determine the net effect estimate Z as a summative or linear combination of effect estimate X and effect estimate Y is based on the same statistical principles for determination of confidence intervals for differences between means, using addition instead of subtraction. That is,

$$Z = X + Y.$$

Assumptions regarding effect estimates include they:

1) represent data conforming to the normal distribution,
2) are independent and not correlated with each other, and
3) are expressed using the same units of measure.
4) The mean (or point estimate) for a net effect estimate Z is simply the addition of the means (or point estimates) for effect estimates X and Y. That is,

$$\text{Mean } Z = \text{Mean } X + \text{Mean } Y.$$

A 95% confidence interval for the net effect estimate is determined by calculating Mean Z +/- 1.96 SD_{MeanZ} where SD_{MeanZ} = standard deviation [SD] of Mean Z.

For the net effect estimate Z, the SD of Mean Z is related to the SDs of the component estimates Mean X and Mean Y through the formula:

$$SD_{MeanZ}^2 = SD_{MeanX}^2 + SD_{MeanY}^2$$

Therefore,

$$SD_{MeanZ} = \sqrt{(SD_{MeanX}^2 + SD_{MeanY}^2)}$$

The 95% confidence interval for the net effect will be:

$$\text{Mean } Z - 1.96\ SD_{MeanZ} \text{ to Mean } Z + 1.96\ SD_{MeanZ}$$

The third assumption (that effect estimates X and Y are expressed using the same units of measure) is rarely true so we need to introduce a “standardization” or “normalization” of outcomes, and this can be

done based on their relative importance. One approach is to assign a multiplier (M) to each outcome representing its importance or relative value compared to a reference outcome. The reference outcome can be external to the body of evidence, or can be one of the outcomes of interest (in which case the value of M for the reference outcome will be 1).

The mean (or point estimate) for a net effect estimate Z, expressed in units of multiples of the reference outcome, becomes:

$$\text{Mean } Z = (M_X \times \text{Mean } X) + (M_Y \times \text{Mean } Y)$$

With the use of multipliers, the SD of the net effect estimate Z becomes related to the formula:

$$SD_{MeanZ}^2 = (M_X \times SD_{MeanX})^2 + (M_Y \times SD_{MeanY})^2$$

Therefore,

$$SD_{MeanZ} = \sqrt{(M_X^2 SD_{MeanX}^2 + M_Y^2 SD_{MeanY}^2)}$$

Note that if the SD_{Mean} is not directly reported for an individual effect estimate, it can be derived from the width of the 95% confidence interval (CIW) for the effect estimate:

$$SD_{MeanX} = CIW_X / 3.92$$

Using the data for the sacubitril-valsartan example (with units of hospitalization-equivalent events per 1000 patients) we get:

$SD_{\text{All-cause mortality outcome}}$ = CIW of 160 / 3.92 = 40.816

$SD_{\text{Hospitalization rate outcome}}$ = CIW of 27 / 3.92 = 6.888

$SD_{\text{Symptomatic hypotension rate outcome}}$ = CIW of 12 / 3.92 = 3.061

Applying $SD_{MeanZ} = \sqrt{(SD_{MeanX}^2 + SD_{MeanY}^2)}$ we get:

$SD_{\text{Net effect estimate}} = \sqrt{(SD_{\text{All-cause mortality}}^2 + SD_{\text{Hospitalization rate}}^2 + SD_{\text{Symptomatic hypotension rate}}^2)}$

$SD_{\text{Net effect estimate}} = \sqrt{(40.816^2 + 6.888^2 + 3.061^2)}$

$SD_{\text{Net effect estimate}} = \sqrt{(1665.9459 + 47.4445 + 9.3697)}$

$SD_{\text{Net effect estimate}} = \sqrt{(1722.7601)} = 41.5$

The 95% confidence interval for the net effect estimate is the mean +- 1.96 SD. For the lower boundary, this translates to 154 - (1.96)(41.5) = 154-81.34 = 72.66 (rounded to 73) and for the upper boundary, this would be 154 + 81.34 = 235.34 (rounded to 235).

We report a net effect estimate of a decrease in 154 hospitalization-equivalent events per 1000 patients (95% confidence interval for the net effect estimate being 73 fewer to 235 fewer hospitalization-equivalent events per 1000 patients).

Sensitivity analysis of the net effect estimate

The 95% confidence interval implies a range within which the true net effect is likely to occur. There are many factors that can affect the certainty that the true net effect is within this range.

If assumptions used in the model are not met, the results will not have accurate precision. If individual outcomes are correlated (such as increase in one benefit being correlated with an increase in another benefit), the "true" 95% confidence interval would be wider or less precise than the one estimated by our method. Alternatively, if individual outcomes are inversely correlated (such as an increase in a benefit being correlated with an increase in a harm, or correlated with a decrease in another benefit), then the "true" 95% confidence interval would be narrower or more precise than the one estimated by our method. In the latter case our proposed approach is conservative but less powerful. If outcomes have other dependencies or do not follow a normal distribution (such as a highly skewed distribution), then the 95% confidence interval may be inaccurate.

The statistical formulas can be adjusted with correlation coefficients if they can be estimated. The formula to determine the standard deviation of the mean of the net effect

$$SD_{MeanZ}^2 = (M_X \times SD_{MeanX})^2 + (M_Y \times SD_{MeanY})^2$$

is modified to

$$SD_{MeanZ}^2 = (M_X \times SD_{MeanX})^2 + (M_Y \times SD_{MeanY})^2 + (2 \times r \times M_X \times SD_{MeanX} \times M_Y \times SD_{MeanY})$$

where r = the correlation coefficient between X and Y. Correlation coefficients are rarely available but the maximum value of r that appears plausible can be used for a sensitivity analysis to address plausible correlations between outcomes.

For the sacubitril-valsartan example, there is data suggesting a small inverse correlation (r = -0.17) between all-cause mortality and hospitalization for heart failure among patients with heart failure (25). There is no data addressing correlations between drug-related symptomatic hypotension and the outcomes of mortality or hospitalization. Let's assume r = 0.5 for each of these as an upper bound of plausible correlations for a sensitivity analysis.

$$SD_{\text{Net effect estimate}} = \sqrt{(SD_{\text{All-cause mortality}}^2 + SD_{\text{Hospitalization rate}}^2 + SD_{\text{Symptomatic hypotension rate}}^2) + {}_{2r}(SD_{\text{Mortalty}})(SD_{\text{Hospiltalization}}) + {}_{2r}(SD_{\text{Mortalty}})(SD_{\text{Symptomatic hypotension rate}}) + {}_{2r}(SD_{\text{Symptomatic hypotension rate}})(SD_{\text{Hospitalization}})}$$

$$SD_{\text{Net effect estimate}} = \sqrt{(40.816^2 + 6.888^2 + 3.061^2) + 2(-0.17)(40.816)(6.888) + 2(0.5)(40.816)(3.061) + 2(0.5)(3.061)(6.888)}$$

$$SD_{\text{Net effect estimate}} = \sqrt{(1722.7601) + (-95.5878) + (124.9378) + (21.0842)}$$

$$SD_{\text{Net effect estimate}} = \sqrt{(1773.1943)} = 42.1$$

This net effect estimate (in a sensitivity analysis adjusting for known and plausible correlations among outcomes) is a decrease in 154 hospitalization-equivalent events per 1000 patients (95% confidence interval 71 fewer to 237 fewer hospitalization.

Example 1. Longer dual-antiplatelet therapy (DAPT) after drug-eluting stents

A systematic review comparing longer versus shorter durations of DAPT after drug eluting stent placement provides the summary of effect estimates for longer duration DAPT in Appendix Table 1 (26). Longer duration of DAPT ranged from 12 months to 42 months and shorter duration of DAPT ranged from 3 months to 18 months (26).

Appendix Table 1. Summary of findings for longer versus shorter durations of dual antiplatelet therapy after drug eluting stent placement

Outcome	Absolute effect estimate per 1000 patients (95% confidence interval)	Certainty of effect estimates
All-cause mortality	2 more (0 change to 4 more)	High*
Myocardial infarction	8 fewer (12 fewer to 2 fewer)	Moderate
Major bleeding	6 more (3 more to 10 more)	High
Any stroke	0 change (2 fewer to 2 more)	High*

* originally reported as moderate quality evidence with downgrade limited to precision. Precision downgrade for single outcome effect estimates are not relevant in this approach as the confidence intervals are being used in the determination of the net effect estimate.

Step 1. Determine the outcomes to be combined.

All four outcomes (mortality, myocardial infarction, major bleeding, and stroke) are considered impactful to include in net effect estimates. None are overlapping outcomes with the assumptions that hemorrhagic stroke contributes minimally to estimates of major bleeding, and fatal outcomes contribute minimally to estimates of myocardial infarction, major bleeding and stroke.

Step 2. Determine the quantified relative importance for each outcome.

Myocardial infarction-equivalent will be considered the reference unit. For example purposes, we will start with the assumption that patients would consider the importance of a myocardial infarction and major bleeding similarly, consider a stroke 3 times more important, and consider mortality 5 times more important. These assignments of relative importance of outcomes are derived from systematic review of evidence of relative importance of outcomes for myocardial infarction, major bleeding and stroke (17) and without empiric investigation for the mortality outcome (3).

Step 3. Combine the importance-adjusted effect estimates.

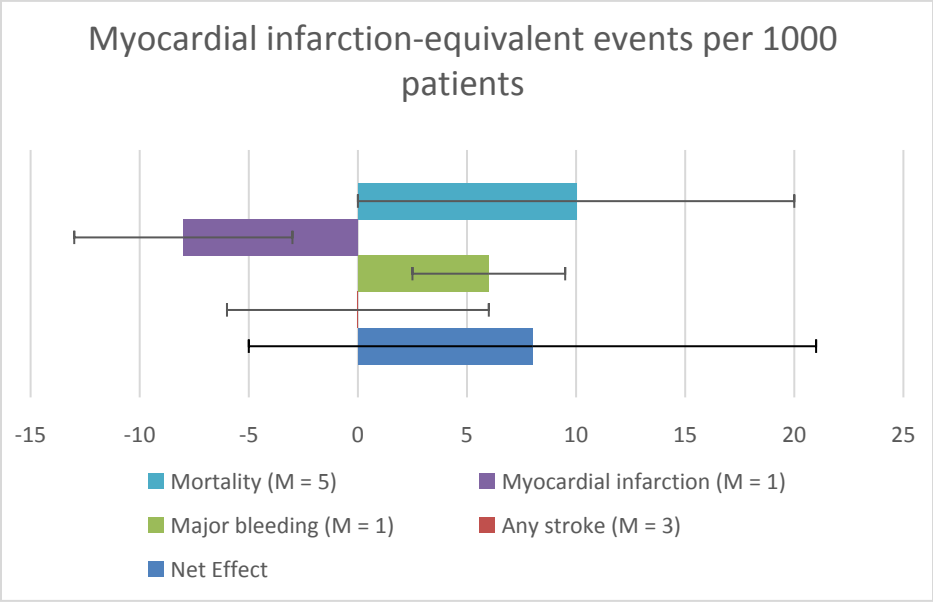
Importance-adjusted effect estimates are determined by multiplying each effect estimate by its relative importance multiplier. Our importance-adjusted effect estimates (in units of myocardial infarction-equivalent events per 1000 patients) are summarized in Appendix Table 2:

Appendix Table 2. Importance-adjusted effect estimates for longer versus shorter durations of dual antiplatelet therapy after drug eluting stent placement

Outcome	Absolute effect estimate (in units of myocardial infarction-equivalent events per 1000 patients) (95% confidence interval)
All-cause mortality	10 more (0 change to 20 more)
Myocardial infarction	8 fewer (12 fewer to 2 fewer)
Major bleeding	6 more (3 more to 10 more)
Any stroke	0 change (6 fewer to 6 more)

The effect estimates are combined using the online calculator at ebscohealth.com/innovations (see Appendix Part 2). The net effect estimate is an increase in 8 myocardial infarction-equivalent events per 1000 patients (95% confidence interval [CI] decrease in 5 to increase in 21 myocardial infarction-equivalent events per 1000 patients).

Appendix Figure 1. Effect estimates for longer versus shorter durations of dual antiplatelet therapy after drug eluting stent placement



Step 4. Classify the precision of the net effect estimate.

The net effect point estimate is harmful, the lower bound of the confidence interval for the net effect estimate is beneficial, and the absolute value of the lower bound of the confidence interval is smaller than the absolute value of the net effect point estimate. This pattern is likely net harm, and consistent with a moderate certainty of net harm.

Step 5. Consider the certainty of effect estimates for outcomes that are critical to the likelihood of net benefit.

Mortality and major bleeding are critical outcomes (potential differentiators of the likelihood of net benefit) because removal of either outcome could change the pattern to one suggesting net benefit. Stroke and myocardial infarction have limited impact on the net effect classification. Both critical outcomes have high certainty of evidence so this does not change our moderate certainty of net harm.

Step 6. Consider the range of relative importance for outcomes. Determine if the net effect estimate across the range of relative importance changes the certainty of net benefit rating.

If patients considered reduction of myocardial infarction to have higher relative importance than mortality and major bleeding it is possible to derive a net benefit. Such relative importance ratings are plausible because myocardial infarction can have a greater contribution to long-term quality of life. Consideration of the range of relative importance for outcomes leads to a low certainty of net harm.

Completing the evidence-to-decision framework

With a low certainty of net harm, the expected result is a weak recommendation against longer duration DAPT after drug-eluting stent placement. The costs are relatively low and there are little adverse consequences related to acceptability, feasibility and equity, so guideline panels may consider to make a weak recommendation against longer duration DAPT.

At the current time, major guidelines have inconsistent recommendations for this concept. The American College of Chest Physicians makes a strong recommendation against DAPT (and for single antiplatelet therapy) after 12 months following drug-eluting stent placement (27). The American College of Cardiology makes a weak recommendation suggesting continuing DAPT beyond 12 months may be considered in patients receiving drug-eluting stents (28).

Example 2. Sacubitril-valsartan for symptomatic heart failure

This example is a decision or recommendation to use sacubitril-valsartan instead of an angiotensin-converting enzyme (ACE) inhibitor or angiotensin receptor blocker (ARB) in patients with symptomatic heart failure with reduced ejection fraction despite treatment with an ACE inhibitor or ARB. A systematic evidence review and GRADE evidence profile for such use of sacubitril-valsartan finds the effect estimates in Appendix Table 3 (29, 30), based on a single trial (31).

Appendix Table 3. Summary of findings for sacubitril-valsartan versus continued ACE inhibitor or ARB in symptomatic heart failure with reduced ejection fraction

Outcome	Absolute effect estimate per 1000 patients (95% confidence interval)	Certainty of evidence*
All-cause mortality	29 fewer (12 fewer to 44 fewer)	Moderate
Cardiovascular mortality	31 fewer (17 fewer to 45 fewer)	Moderate
Hospitalization for worsening heart failure	31 fewer (16 fewer to 43 fewer)	Moderate
Symptomatic hypotension	44 more (33 more to 57 more)	Moderate
Change in heart failure symptom score (scale 0-100)	1.64 points decrease (0.63-point decrease to 2.65-point decrease)	Moderate
Decline in renal function	4 fewer (3 fewer to 9 fewer)	Moderate

* Certainty of evidence ratings here do not rate down for imprecision.

Step 1. Determine the outcomes to be combined

Two outcomes were dropped from consideration because they were considered to have little to no impact on the net effect. In the sacubitril-valsartan example decline in renal function was considered not impactful for determination of the net effect estimate because the effect size is small and the outcome has low importance to patients. Change in heart failure symptom score was considered not impactful for determination of the net effect estimate because the effect size is small and the relative importance is uncertain and may be accounted for in other outcomes. Using means for a continuous score can be misleading when one considers the impact on individual patients who vary in their responses (i.e. assuming every patient experiences the mean effect is likely an erroneous assumption). The only data regarding the proportion of patients who have an important change in symptoms is the outcome of hospitalization for worsening heart failure, and authors of the study reported this outcome.

All-cause mortality is selected instead of cardiovascular mortality to avoid duplicate counting of mortality. The outcomes included in net effect estimation are all-cause mortality, hospitalization for worsening heart failure and symptomatic hypotension.

Step 2. Determine the quantified relative importance for each outcome

Hospitalization-equivalent events per 1000 patients will be considered the reference unit. We do not readily find empiric evidence for the relative importance of outcomes in patients with heart failure. We

will start with the assumption that patients would consider the outcome of all-cause mortality 5 times more important than an episode of hospitalization, and an outcome of symptomatic hypotension half as important as being hospitalized.

Step 3. Combine the importance-adjusted effect estimates

Importance-adjusted effect estimates are determined by multiplying each effect estimate by its relative importance multiplier. Our importance-adjusted effect estimates (in units of hospitalization-equivalent events per 1000 patients) are summarized in Appendix Table 4:

Appendix Table 4. Importance-adjusted effect estimates for sacubitril-valsartan versus continued ACE inhibitor or ARB in symptomatic heart failure with reduced ejection

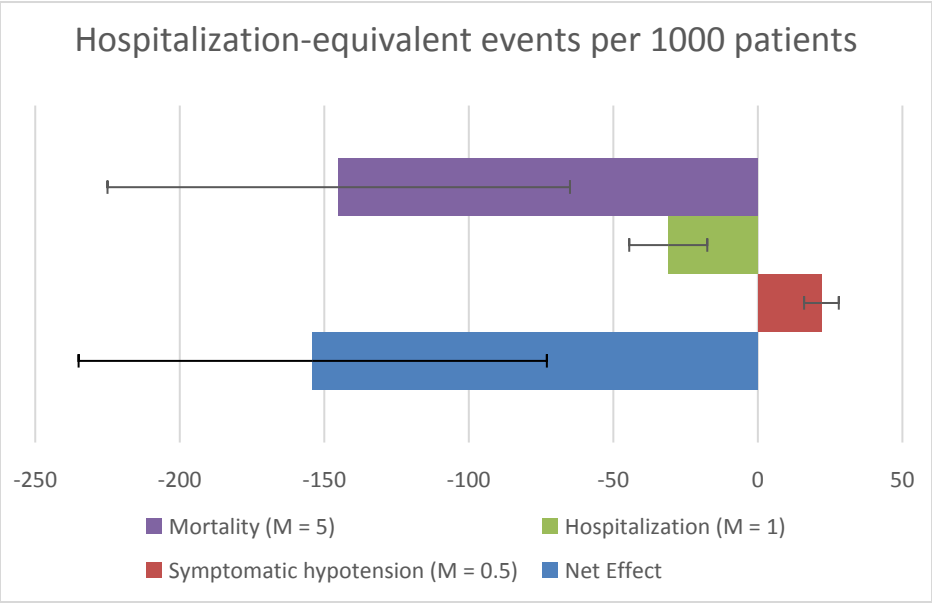
Outcome	Absolute effect estimate (in units of hospitalization-equivalent events per 1000 patients) (95% confidence interval)
All-cause mortality	145 fewer (60 fewer to 220 fewer)
Hospitalization for heart failure	31 fewer (16 fewer to 43 fewer)
Symptomatic hypotension	22 more (16.5 more to 28.5 more)

The effect estimates are combined using the online calculator at ebscohealth.com/innovations and the calculations are shown in part in Appendix Part 2.

The net effect point estimate is a decrease in 154 hospitalization-equivalent events per 1000 patients. (-145 plus -31 plus +22 = -154)

The net effect estimate is a decrease in 154 hospitalization-equivalent events per 1000 patients. (95% CI 73 fewer to 235 fewer hospitalization-equivalent events per 1000 patients)

Appendix Figure 2. Effect estimates for sacubitril-valsartan versus continued ACE inhibitor or ARB in symptomatic heart failure with reduced ejection fraction



Step 4. Classify the precision of the net effect estimate.

The entire confidence interval of the net effect estimate is beneficial so the pattern is net benefit, consistent with a high certainty of net benefit.

Step 5. Consider the certainty of effect estimates for outcomes that are critical to the likelihood of net benefit.

Mortality is potentially differentiating because removal of a mortality effect would change the net effect estimate from 154 fewer (95% CI 73 fewer to 235 fewer) hospitalization-equivalent events per 1000 patients to 9 fewer (95% CI 24 fewer to 6 more) events per 1000 patients, and the overall pattern would change from net benefit to likely net benefit.

Hospitalization for heart failure is not potentially differentiating because removal from the net effect estimate would not change the pattern from net benefit. The net effect estimate would be 123 fewer (95% CI 43 fewer to 203 fewer) events per 1000 patients. A result of increasing hospitalizations for heart failure is not a plausible likelihood. One could question whether total hospitalizations should be used as an outcome rather than cause-specific hospitalization. The outcome of total hospitalizations was not reported in the underlying evidence (28), and guideline panels would need to determine if such an outcome is impactful enough to reassess the overall balance of benefits and harms for this decision.

Symptomatic hypotension is initially not potentially differentiating because removal from the net effect estimate would not change the pattern from net benefit. Symptomatic hypotension can still be considered critical because a higher rate of symptomatic hypotension than observed in the underlying

evidence is plausible, especially related to the use of run-in periods excluding patients who did not tolerate study medications.

The critical outcomes have effect estimates with moderate certainty. This leads to a moderate certainty of net benefit.

Step 6. Consider the range of relative importance for outcomes. Determine if the net effect estimate across the range of relative importance changes the certainty of net benefit rating.

We started with an assumption that the average patient would consider the importance of all-cause mortality five times more important than an episode of hospitalization, and an outcome of symptomatic hypotension half as important as being hospitalized. To consider a range of relative importance for outcomes we should consider the lowest relative importance for all-cause mortality and highest relative importance for symptomatic hypotension that would occur among patients facing this decision and is considered reasonable to reflect the range of importance among common, rational people. Some patients (such as those with terminal illness) may place higher importance on how they feel than mortality so for these patients they might consider mortality, symptomatic hypotension, and hospitalization to be equivalent.

Using assumptions of equivalence across these three outcomes the net effect estimate would be 16 fewer (95% CI 40 fewer to 9 more) events per 1000 patients.

With a reasonable limit for the range of relative importance (including most patients) weighted to support net harm, the net effect estimate changes from net benefit to likely net benefit. If there were otherwise high certainty of net benefit this finding could reduce our certainty to moderate certainty of net benefit. As we already have a moderate certainty of net benefit, extreme assumptions reaching likely net benefit do not further change our certainty.

Completing the evidence-to-decision framework

In an assessment in 2015 the moderate certainty of net benefit justified a weak recommendation for sacubitril-valsartan (29, 30). The high cost further supported a weak recommendation. Four national guidelines have since made strong recommendations for the use of sacubitril-valsartan (32-35), though the findings have not been replicated in a second trial. A recommendations panel reconsidered the rationale across all four guidelines and reconfirmed a weak recommendation for sacubitril-valsartan based on a moderate certainty of evidence (limited to a single trial with potential selection bias related to the run-in period), a moderate certainty of net benefit (considering the range of quantitative estimates of importance of outcomes), and high cost with some uncertainty in the cost-benefit ratio (29). A different recommendations panel could generate different ratings, but the process allows explicit and transparent expression of what is being rated and how it is rated.

Example 3. Ivabradine for symptomatic heart failure

Ivabradine is a heart rate lowering drug which has been tested for clinical use in patients with heart failure in two large randomized trials (36, 37). In the first trial ivabradine was not associated with overall clinical benefit and was not associated with any decrease in death or hospitalization attributed to heart failure (36). In the second trial with more stringent selection criteria (left ventricular ejection fraction <= 35%, heart rate => 70 beats/minute) ivabradine reduced the rate of hospital admissions for worsening heart failure (37).

Outcome differences with ivabradine instead of placebo (from randomization until first event, up to 42 months) are summarized in Appendix Table 5 (38):

Appendix Table 5. Summary of findings for ivabradine versus placebo in symptomatic heart failure with reduced ejection fraction

Outcome	Absolute effect estimate per 1000 patients (95% confidence interval)	Certainty of effect estimates
All-cause mortality	13.9 fewer (31.8 fewer to 4 more)	Moderate
Cardiovascular mortality	11.9 fewer (29 fewer to 5.2 more)	Moderate
Death from heart failure	11.4 fewer (21 fewer to 1.8 fewer)	Moderate
Hospitalization for any cause	35.6 fewer (59.4 fewer to 11.8 more)	Moderate
Hospitalization for cardiovascular reason	42.3 fewer (65 fewer to 19.6 more)	Moderate
Hospitalization for worsening heart failure	47.3 fewer (66 fewer to 28.6 fewer)	Moderate
Bradycardia	33.5 more (26 more to 41 more)	High
Phosphenes (a visual adverse effect)	22.6 more (16.5 more to 28.8 more)	High
Atrial fibrillation	12.1 more (2.1 more to 22 more)	High

The second trial had a low risk of bias though the quality of evidence could be considered moderate for benefits based on inconsistency with the first trial. The adverse effects data could be considered as high quality evidence as the findings are consistent with the first trial (39).

Step 1. Determine the outcomes to be combined.

All-cause mortality, hospitalization for any cause, bradycardia, phosphenes, and atrial fibrillation are selected as non-overlapping outcomes.

Step 2. Determine the quantified relative importance for each outcome.

Hospitalization-equivalent relative importance will be estimated at 0.3 for each adverse effect and 5 for mortality.

Step 3. Combine the importance-adjusted effect estimates

The importance-adjusted effect estimates (in units of hospitalization-equivalent events per 1000 patients) are in Appendix Table 6.

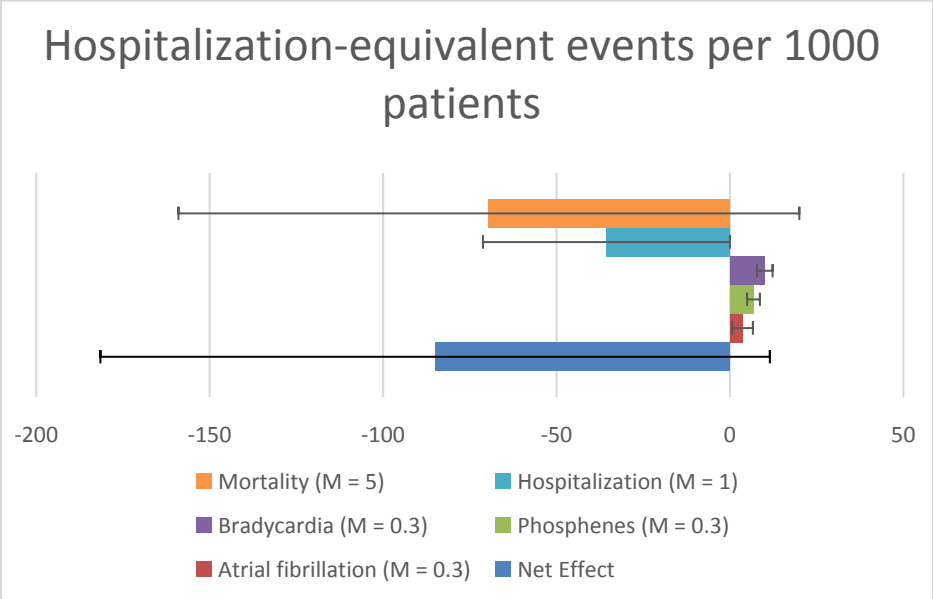
Appendix Table 6. Importance-adjusted effect estimates for ivabradine versus placebo in symptomatic heart failure with reduced ejection fraction

Outcome	**Absolute effect estimate** (in units of hospitalization-equivalent events per 1000 patients) **(95% confidence interval)**	**Confidence interval width (CIW)**
All-cause mortality	69.5 fewer (159 fewer to 20 more)	179 per 1000
Hospitalization for any cause	35.6 fewer (59.4 fewer to 11.8 more)	71.2 per 1000
Bradycardia	10.05 more (7.8 more to 12.3 more)	4.5 per 1000
Phosphenes (a visual adverse effect)	6.78 more (4.95 more to 8.64 more)	3.69 per 1000
Atrial fibrillation	3.63 more (0.63 more to 6.6 more)	5.97 per 1000

The net effect point estimate is a decrease in 85 hospitalization-equivalent events per 1000 patients. (-69.5 plus -35.6 plus +10.05 plus +6.78 plus +3.63 = -84.64)

The net effect estimate is a decrease in 85 (95% confidence interval decrease in 181 to increase in 12) hospitalization-equivalent events per 1000 patients (see Appendix Figure 3).

Appendix Figure 3. Effect estimates for ivabradine versus placebo in symptomatic heart failure with reduced ejection fraction



Step 4. Classify the precision of the net effect estimate.

The pattern is likely net benefit, consistent with a moderate certainty of net benefit.

Step 5. Consider the certainty of effect estimates for outcomes that are critical to the likelihood of net benefit.

As the effect estimates have at least moderate certainty of evidence there is moderate certainty of net benefit.

Step 6. Consider the range of relative importance for outcomes. Determine if the net effect estimate across the range of relative importance changes the certainty of net benefit rating.

A plausible range of relative importance could include consideration of death equivalent to hospitalization and other adverse effects 0.6 times as disruptive as hospitalization. Such assumptions would lead to importance-adjusted effect estimates of:

- All-cause mortality: 13.9 fewer (95% CI 31.8 fewer to 4 more)
- Hospitalization for any cause: 35.6 fewer (95% CI 59.4 fewer to 11.8 more)
- Bradycardia: 20.1 more (95% CI 15.6 more to 24.6 more)
- Phosphenes (a visual adverse effect): 13.56 more (95% CI 9.90 more to 17.28 more)
- Atrial fibrillation: 7.26 more (95% CI 1.26 more to 13.2 more)

These estimates would result in a net effect estimate of a decrease in 9 (95% confidence interval decrease in 49 to increase in 32) hospitalization-equivalent events per 1000 patients.

This would be possible net benefit, and results in a low certainty of net benefit upon consideration across the range of relative importance that patients may have for the various effects.

Completing the Evidence-to-Decision Framework

A low certainty of net benefit supports a weak recommendation for ivabradine in patients meeting the selected criteria used in the trial suggesting benefit. Three of four current guidelines provide a weak recommendation for ivabradine in this setting (40-42) while one makes a strong recommendation (43).

Example 4. Second autologous stem cell transplant (ASCT) for patients with relapsed myeloma and response duration more than 2 years after first ASCT

A National Institute for Health and Clinical Excellence (NICE) guideline includes GRADE profiles for a second ASCT in relapsed myeloma including (44):

- Median overall survival from relapse – low quality evidence – absolute effect 2.1 years longer (95% CI not reported)
- Median time to progression – moderate quality evidence – absolute effect 13 months longer (95% CI not reported)
- No evidence identified for treatment-related morbidity and mortality, health-related quality of life, and adverse effects.

Step 1. Determine the outcomes to be combined.

The guideline panel considered overall survival and progression-free survival to be the most impactful outcomes for consideration. Because overall survival includes progression-free survival, progression events (time to progression) and death events (time to death) can be counted as non-overlapping outcomes.

Step 2. Determine the quantified relative importance for each outcome.

Time to death (overall survival) will be considered the reference unit. We will start with the assumption that patients would consider time to progression 0.2 times the importance of time to death.

Step 3. Combine the importance-adjusted effect estimates.

The importance-adjusted estimate for time to progression is median 2.6 months (0.2 x 13 months) and the estimate for time to death is median 2.1 years (or 25.2 months).

The point estimate for the net effect is the equivalent of 27.8 months of increased survival. Confidence intervals were not reported.

Step 4. Classify the precision of the net effect estimate.

This appears to start with a pattern of net benefit. The statistical significance was not expressed but assuming the results were statistically significant the confidence intervals would be completely within estimates of benefit because no evidence was provided to suggest harm.

Step 5. Consider the certainty of effect estimates for outcomes that are critical to the likelihood of net benefit.

Overall survival is the critical outcome here and was reported as low certainty of evidence based on a single retrospective comparative study (and related consistent data in noncomparative studies). Even so the guideline panel could potentially consider this to represent a moderate certainty of a survival

benefit and a low certainty for a specific magnitude of effect. This could lead to a moderate certainty of net benefit.

Step 6. Consider the range of relative importance for outcomes. Determine if the net effect estimate across the range of relative importance changes the certainty of net benefit rating.

In a model with no harms being considered as important outcomes, the range of relative importance is really an opportunity to consider how potential harms may affect the balance of benefits and harms. The guideline panel rationalized that harms would be similar to what patients experienced with their first ASCT and patients would thus have individual experience representing their individual harms estimates when considering the balance of harms and benefits.

Completing the Evidence-to-Decision Framework

In the context of harms mainly being considered burdens the patient would individually consider, the potential for increases in overall survival is considered a moderate certainty of net benefit. To reflect the importance of the patient weighing a personalized relative importance the guideline panel made a strong recommendation to offer the therapy (for the potential for net benefit) rather than recommend that the therapy should be administered.

Example 5. Avoiding 100% oxygen saturation in intensive care unit

A randomized trial with 480 adults admitted to the intensive care unit (ICU) found mortality in the ICU of 11.6% with target arterial oxyhemoglobin saturation (SpO2) 94%-98% versus 20.2% with target SpO2 97%-100% (absolute risk reduction 8.6%, 95% CI 1.7% to 15%) (45). This evidence may be considered to have moderate certainty due to early trial termination without use of a formal stopping rule.

Step 1. Determine the outcomes to be combined.

Although other outcomes were reported, there was no evidence of benefits for high-SpO2, so the mortality outcome can be considered the primary outcome for a net effect estimate.

Step 2. Determine the quantified relative importance for each outcome.

Step 3. Combine the importance-adjusted effect estimates.

These steps are irrelevant in this case and the net effect estimate is the estimate for ICU mortality, which can be considered inversely for the action of targeting an SpO2 97%-100% (absolute risk increase 8.6%, 95% confidence interval 1.7% to 15%).

Step 4. Classify the precision of the net effect estimate.

There is a net harm based on the confidence intervals of the effect estimate, consistent with a high certainty of net harm.

Step 5. Consider the certainty of effect estimates for outcomes that are critical to the likelihood of net benefit.

As the underlying evidence is a single trial with early unplanned termination, the moderate certainty of evidence may reduce the certainty of net harm to moderate.

Step 6. Consider the range of relative importance for outcomes. Determine if the net effect estimate across the range of relative importance changes the certainty of net benefit rating.

This is irrelevant following the decision to focus on a single outcome

Completing the Evidence-to-Decision Framework

A moderate certainty of net harm is sufficient to support a strong recommendation against an intervention with no apparent benefit.

Summary of Examples.

Five examples are presented to show how the model for defining and reporting the certainty of net benefit or certainty of net harm can provide more clear and explicit representations of evidence-based assessments and judgments supporting a recommendation spanning a broad continuum of complex situations. See Appendix Table 7.

Example 1 (Longer DAPT after drug-eluting stents) shows a net effect estimate suggesting a low certainty of net harm. Adjustment for certainty of evidence and the range of relative importance across outcomes is unnecessary as the certainty is already low. No other factors change the approach to the recommendation and we support a weak recommendation against. This may be clearer than the variations across current guidelines ranging from a weak recommendation for to a strong recommendation against.

Example 2 (Sacubitril-valsartan for symptomatic heart failure on standard therapy) is an example in which a net effect estimate shows net benefit based on a single trial using reasonable assumptions of relative importance of outcomes. The moderate certainty of evidence and the influence of reasonable extremes of relative importance assignments each led to ratings of a moderate certainty of net benefit. A moderate certainty of net benefit and high cost may support a weak recommendation for sacubitril-valsartan although many current guidelines provide a strong recommendation.

Example 3 (Ivabradine for symptomatic heart failure) shows a treatment with relatively smaller effects on benefits and harms with a closer balance between benefits and harms and moderate certainty of effect estimates. The resulting low certainty of net benefit supports a weak recommendation for ivabradine, and most current guidelines provide a weak recommendation for it.

Example 4 (Second ASCT for patients with relapsed myeloma and response duration more than 2 years after first ASCT) starts with low to moderate certainty in effect estimates for benefits and no direct comparative evidence to quantify harms. This leads to a higher certainty of net benefit, though still a moderate certainty of net benefit given the limited certainty in effect estimates. Despite not reaching a high certainty of net benefit the guideline panel reasoned that the harms for a second ASCT would be patient-specific (and patient-recognized based on the first ASCT) so provided a strong recommendation to offer the therapy and allow the patient to individually weigh the estimated benefits against their individualized harms. This is consistent with the GRADE approach which would provide a weak recommendation and encourage shared decision making. The guideline panel did not provide a strong recommendation for the intervention without shared decision making.

Example 5 (Avoiding 100% oxygen saturation in intensive care unit) is an example of an intervention with no apparent benefit and moderate certainty of net harm. The quantitative effect estimates support a high certainty of net harm but the risk of bias (qualitative certainty) reduced the overall assessment to a moderate certainty of net harm. Even so, without any apparent benefit, a strong recommendation against would be justified.

Appendix Table 7. Certainty of net benefit and strength of recommendations in examples

Example	Certainty of Evidence for Critical Outcomes*	Certainty of Net Benefit	Strength of Recommendation
Longer dual-antiplatelet therapy (DAPT) after drug-eluting stents	Moderate to high certainty of evidence	Low certainty of net harm	Weak recommendation against
Sacubitril-valsartan for symptomatic heart failure on standard therapy	Moderate certainty of evidence	Moderate certainty of net benefit	Weak recommendation for
Ivabradine for symptomatic heart failure	Moderate certainty of evidence	Low certainty of net benefit	Weak recommendation for
Second autologous stem cell transplant (ASCT) for patients with relapsed myeloma and response duration more than 2 years after first ASCT	Low to moderate certainty of evidence	Moderate certainty of net benefit	Weak recommendation for (or strong recommendation for offering with shared decision making)
Avoiding 100% oxygen saturation in intensive care unit	Moderate certainty of evidence	Moderate certainty of net harm	Strong recommendation against

* Certainty of evidence ratings here do not rate down for imprecision.

References

References 1-24 are the same as those included in the index article.

1. URL: www.gradeworkinggroup.org. Accessed November 27, 2016.
2. Atkins D, Best D, Briss PA, et al; GRADE Working Group. Grading quality of evidence and strength of recommendations. BMJ 2004;328:1490. doi: 10.1136/bmj.328.7454.1490.
3. Balshem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;64:401-6. doi: 10.1016/j.jclinepi.2010.07.015.
4. Hultcrantz M, Rind D, Akl E, et al; GRADE Working Group. The GRADE Working Group clarifies the construction of certainty of evidence. J Clin Epidemiol 2017;87:4-13. doi: 10.1016/j.jclinepi.2017.05.006.
5. Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. J Clin Epidemiol 2013;66:726-35. doi: 10.1016/j.jclinepi.2013.02.003.
6. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.3.2 Confidence in best estimates of magnitude of effects (quality of evidence). URL: gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.s1fsemn2yjpm. Accessed November 27, 2016.
7. Guyatt GH, Oxman AD, Kunz R, et al; GRADE Working Group. Going from evidence to recommendations. BMJ 2008;336:1049-51. doi: 10.1136/bmj.39493.646875.AE.
8. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.3.3 Confidence in values and preferences. URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.i5hfweocv3qs. Accessed November 27, 2016.
9. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.2.1 Balance of desirable and undesirable consequences. URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.feb5g22tf6nv. Accessed November 27, 2016.
10. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.3.4 Resource use (cost). URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.y0ey8465h47v. Accessed November 27, 2016.
11. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 6.5 The Evidence-to-Decision framework. URL: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.xr5ac2p2khuq. Accessed November 27, 2016.
12. Sawaya GF, Guirguis-Blake J, LeFevre M, Harris R, Petitti D, U.S. Preventive Services Task Force. Update on the methods of the U.S. Preventive Services Task Force: estimating certainty and

magnitude of net benefit. Ann Intern Med 2007;147:871-5. doi: 10.7326/0003-4819-147-12-200712180-00007.

13. Krist AH, Wolff TA, Jonas DE, et al. Update on the methods of the U.S. Preventive Services Task Force: methods for understanding certainty and net benefit when making recommendations. Am J Prev Med 2018;54:S11-S18. doi: 10.1016/j.amepre.2017.09.011.
14. Editors: Schunemann H, Brozek J, Guyatt G, Oxman A. GRADE Handbook. Handbook for grading the quality of evidence and the strength of recommendations using the GRADE approach. Updated October 2013. Section 3 Selecting and rating the importance of outcomes. URL: https://gdt.gradepro.org/app/handbook/handbook.html#h.1i2bwkm8zpjo. Accessed April 6, 2019.
15. Hazlewood GS, Bombardier C, Tomlinson G, Marshall D. A Bayesian model that jointly considers comparative effectiveness research and patients' preferences may help inform GRADE recommendations: an application to rheumatoid arthritis treatment recommendations. J Clin Epidemiol 2018;93:56-65. doi: 10.1016/j.jclinepi.2017.10.003.
16. Fain KM, Yu T, Li T, et al. Evidence selection for a prescription drug's benefit-harm assessment: challenges and recommendations. J Clin Epidemiol 2016;74:151-7.
17. MacLean S, Mulla S, Akl EA, et al. Patient Values and Preferences in Decision Making for Antithrombotic Therapy: A Systematic Review: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 2012;141:e1S-e23S.
18. Richards T, Schroter S, Price A, Godlee F. Better together: patient partnership in medical journals. BMJ 2018;362:k3798.
19. Schünemann HJ. Interpreting GRADES's levels of certainty or quality of the evidence: GRADE for statisticians, considering review information size or less emphasis on imprecision? J Clin Epidemiol 2016;75:6-15.
20. Efron B, Tibshirani RJ. An Introduction to the Bootstrap, 1st Edition Chapman and Hall/CRC. New York, NY 1993
21. Spiegelhalter DJ. Myles JP, Jones DR, Abrams KR. An introduction to Bayesian methods in health technology assessment. BMJ 1999;319: 508–512.
22. Zhang Y, Alonso Coello P, et al. GRADE Guidelines: 19. Assessing the certainty of evidence in the importance of outcomes or values and preferences -- Risk of bias and indirectness. J Clin Epidemiol 2018.
23. Smith CM. Origin and uses of primum non nocere—above all, do no harm! J Clin Pharmacol 2005;45:371-7. doi: 10.1177/0091270004273680.
24. Schunemann HJ. Guidelines 2.0: do no net harm-the future of practice guideline development in asthma and other diseases. Current allergy and asthma reports 2011;11:261-8.
25. Krumholz HM, Lin Z, Keenan PS, et al. Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. JAMA 2013;309:587-93.
26. Spencer FA, Prasad M, Vandvik PO, et al. Longer versus shorter duration dual-antiplatelet therapy after drug-eluting stent placement: a systematic review and meta-analysis. Ann Intern Med 2015;163:118-26. doi: 10.7326/M15-0083.
27. Vandvik PO, Lincoff AM, Gore JM, et al. Primary and secondary prevention of cardiovascular disease: Antithrombotic Therapy and Prevention of Thrombosis, 9th ed: American College of Chest Physicians Evidence-Based Clinical Practice Guidelines. Chest 2012;141:e637S-e666S. doi: 10.1378/chest.11-2306.
28. Levine GN, Bates ER, Blankenship JC, et al; American College of Cardiology Foundation; American Heart Association Task Force on Practice Guidelines; Society for Cardiovascular

Angiography and Interventions. 2011 ACCF/AHA/SCAI Guideline for Percutaneous Coronary Intervention. A report of the American College of Cardiology Foundation/American Heart Association Task Force on Practice Guidelines and the Society for Cardiovascular Angiography and Interventions. J Am Coll Cardiol 2011;58:e44-122. doi: 10.1016/j.jacc.2011.08.007.

29. DynaMed Plus [Internet]. Ipswich (MA): EBSCO Information Services. 1995-2016. Record No. 114099, Heart failure with reduced ejection fraction; [updated 2016 Oct 07, cited 2016 Nov 27]; [about 101 screens]. Available from http://www.dynamed.com/topics/dmp~AN~T114099/Heart-failure-with-reduced-ejection-fraction#Sacubitril-valsartan--ARNI-.
30. Database of Evidence Profiles [Internet]. Alper B, Colpitts C, Lang E, et al. Should sacubitril-valsartan vs. ACE inhibitor or ARB be used for symptomatic patients with heart failure and reduced ejection fraction? [updated 2015 Nov 17; cited 2016 Nov 27]. Available from http://dbep.gradepro.org/profile/1FEB6C9D-1F7A-EAE1-AEC5-564B423DF770.
31. McMurray JJ, Packer M, Desai AS, et al; PARADIGM-HF Investigators and Committees. Angiotensin-neprilysin inhibition versus enalapril in heart failure. N Engl J Med 2014;371:993-1004. doi: 10.1056/NEJMoa1409077.
32. WRITING COMMITTEE MEMBERS; Yancy CW, Jessup M, Bozkurt B, et al. 2016 ACC/AHA/HFSA Focused Update on New Pharmacologic Therapy for Heart Failure: An Update of the 2013 ACCF/AHA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. Circulation 2016;134:e282-93. doi: 10.1161/CIR.0000000000000435.
33. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al; Authors/Task Force Members; Document Reviewers. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur J Heart Fail 2016;18:891-975. doi: 10.1002/ejhf.592.
34. Scottish Intercollegiate Guidelines Network (SIGN). National clinical guideline for management of chronic heart failure. SIGN 2016 Mar. SIGN publication no. 147.
35. National Institute for Health and Care Excellence (NICE). Sacubitril valsartan for treating symptomatic chronic heart failure with reduced ejection fraction. Technology appraisal guidance. Published: 27 April 2016. TA388.
36. Fox K, Ford I, Steg PG, Tendera M, Ferrari R; BEAUTIFUL Investigators. Ivabradine for patients with stable coronary artery disease and left-ventricular systolic dysfunction (BEAUTIFUL): a randomised, double-blind, placebo-controlled trial. Lancet 2008;372:807-16. doi: 10.1016/S0140-6736(08)61170-8.
37. Swedberg K, Komajda M, Bohm M, et al; SHIFT Investigators. Ivabradine and outcomes in chronic heart failure (SHIFT): a randomised placebo-controlled study. Lancet 2010;376:875-85. doi: 10.1016/S0140-6736(10)61198-1.
38. ClinicalTrials.gov. Effects of Ivabradine on Cardiovascular Events in Patients With Moderate to Severe Chronic Heart Failure and Left Ventricular Systolic Dysfunction. A Three-year International Multicentre Study (SHIFT). NCT02441218. Last Updated May 9, 2016.
39. ClinicalTrials.gov. The BEAUTIFUL Study: Effects of Ivabradine in Patients With Stable Coronary Artery Disease and Left Ventricular Systolic Dysfunction. NCT00143507. Last Updated June 22, 2015.
40. Writing Committee Members, Yancy CW, Jessup M, Bozkurt B, Butler J, Casey DE Jr, et al. 2016 ACC/AHA/HFSA Focused Update on New Pharmacological Therapy for Heart Failure: An Update

of the 2013 ACCF/AHA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines and the Heart Failure Society of America. Circulation 2016;134:e282-93. doi: 10.1161/CIR. 0000000000000435.

41. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JG, Coats AJ, et al; Authors/Task Force Members. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: The Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. Eur Heart J 2016;37:2129-200. doi: 10.1093/eurheartj/ehw128.
42. National Institute for Health and Care Excellence (NICE). Ivabradine for treating chronic heart failure. Technology appraisal guidance. Published: 28 November 2012. TA267.
43. Scottish Intercollegiate Guidelines Network (SIGN). Management of chronic heart failure: A national clinical guideline. Published: March 2016. SIGN publication no. 147. http://www.sign.ac.uk/assets/sign147.pdf
44. National Collaborating Centre for Cancer. Myeloma: diagnosis and management. NICE Guideline 35, Full guideline, February 2016. https://www.nice.org.uk/guidance/ng35/evidence/full-guideline-2306487277
45. Giardis M, Busani S, Damiani E, et al. Effect of conservative vs conventional oxygen therapy on mortality among patients in an intensive care unit: The Oxygen-ICU Randomized Clinical Trial. JAMA 2016;316:1583-1589. doi: 10.1001/jama.2016.11993.