BMJ Open

Defining certainty of net benefit: a GRADE concept paper

Brian S Alper,[1,2] Peter Oettgen,[1,3] Ilkka Kunnamo,[4,5] Alfonso Iorio,[6] Mohammed Toseef Ansari,[7] M Hassan Murad,[8] Joerg J Meerpohl,[9,10] Amir Qaseem,[11] Monica Hultcrantz,[12,13] Holger J Schünemann,[14] Gordon Guyatt,[14] on behalf of The GRADE Working Group

To cite: Alper BS, Oettgen P, Kunnamo I, *et al*. Defining certainty of net benefit: a GRADE concept paper. *BMJ Open* 2019;**9**:e027445. doi:10.1136/bmjopen-2018-027445

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2018-027445).

Received 23 October 2018
Revised 18 April 2019
Accepted 16 May 2019

For numbered affiliations see end of article.

Correspondence to
Dr Brian S Alper;
balper@ebsco.com

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 favourable. 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 unfavourable. 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 100 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 (eg, using fresh-frozen plasma or vitamin K in a patient receiving warfarin with elevated international normalised ratio and an intracranial bleed).

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

Guideline panels using fully contextualised 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

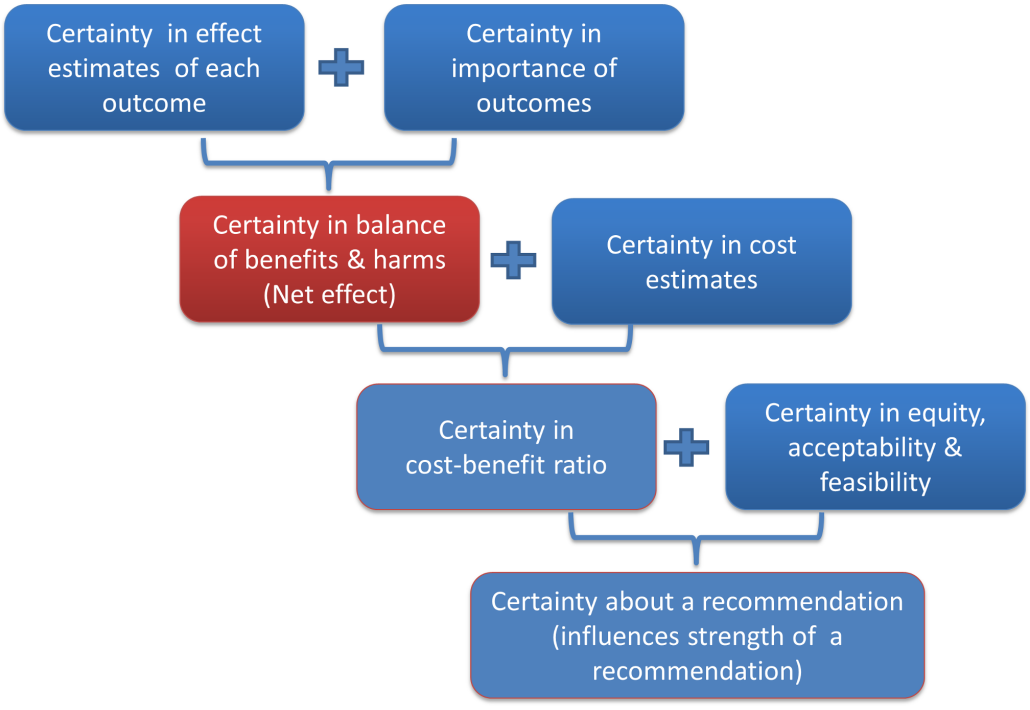
GRADE evidence-to-decision frameworks explicitly identify the following considerations in determining the direction and strength of recommendations:

- 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 healthcare 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 and 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



* 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*.

and contexts, and population-based ratings may be of less interest to patients and clinicians working together to make individual healthcare 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 US Preventive Services Task Force 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 modelling 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 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 are 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 modeller. Simple models can generate confidence intervals (CIs) for a net effect estimate (a composite of individual effect estimates) given the following assumptions (described further in the online supplementary appendix):

1. Effect estimates represent data conforming to normal distributions.

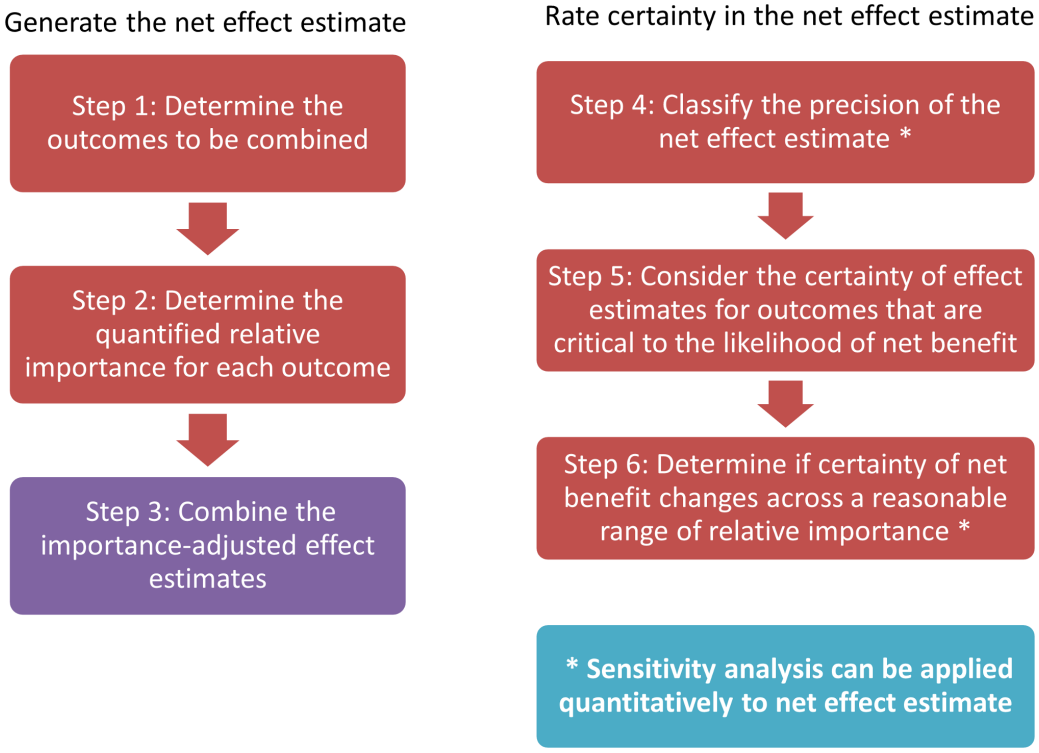


Figure 2 A stepwise approach to determining the certainty of the net effect estimate.

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 favourable 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 online supplementary 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 (eg, all-cause mortality) or to use only the component outcomes (eg, 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 (ie, 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 individualised 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 CIs 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 online supplementary appendix).

Table 1 Classification of precision of net effect estimate

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

*Differentiation of wide versus narrow CIs could be based on a threshold of minimally important differences.

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-cm 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 CI 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 CI 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 CI with both benefit and harm, the lower the certainty in the net benefit. Table 1 and figure 3 present initial suggestions for how these judgements may be made.

The calculation for CIs 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 CIs would be wider or less precise; if inversely correlated, the accurate CIs would be narrower or more precise. If such accuracy is needed, one could add correlation coefficients to the calculation (see online supplementary 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 judgement 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.

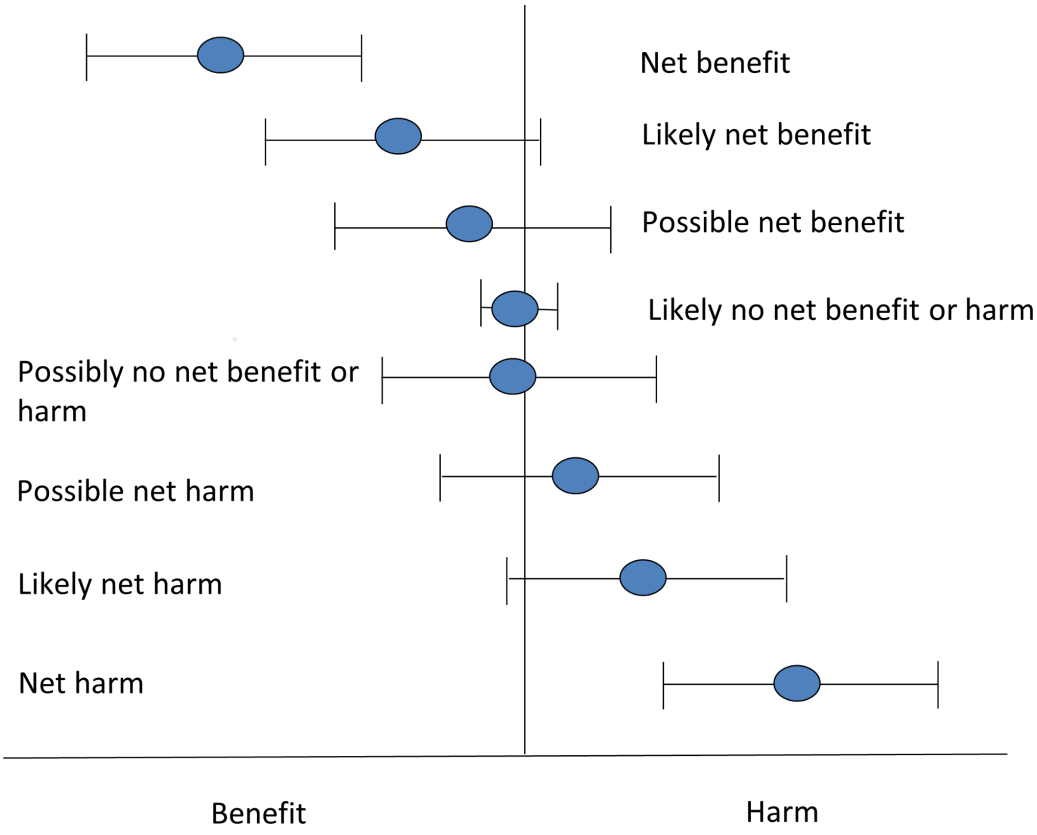


Figure 3 Classification of precision of net effect estimate.

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 (eg, 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 favourable 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 judgements that are already made explicitly or implicitly when guideline panels make recommendations. Reporting the judgements 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 for 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.

Author affiliations
[1]EBSCO Health, DynaMed Plus, EBSCO Information Services Inc., Ipswich, Massachusetts, USA
[2]Department of Family and Community Medicine, University of Missouri-Columbia School of Medicine, Columbia, Missouri, USA
[3]Division of Cardiology, Beth Israel Deaconess Medical Center, Boston, Massachusetts, USA
[4]Duodecim Medical Publications Ltd., Helsinki, Finland
[5]The Finnish Medical Society, Helsinki, Finland
[6]Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada
[7]School of Epidemiology, Public Health and Preventive Medicine, Faculty of Medicine, University of Ottawa, Ottawa, Ontario, Canada
[8]Evidence-Based Practice Center, Mayo Clinic, Rochester, Minnesota, USA
[9]Institute for Evidence in Medicine, Medical Center - University of Freiburg, Freiburg, Germany
[10]Cochrane Germany, Cochrane Germany Foundation, Freiburg, Germany
[11]Department of Clinical Policy, American College of Physicians, Philadelphia, Pennsylvania, USA
[12]Karolinska Institute, Stockholm, Sweden
[13]Statens beredning for medicinsk utvardering, Stockholm, Sweden
[14]Departments of Health Research Methods, Evidence, and Impact and of Medicine, McMaster University Health Sciences Centre, Hamilton, Ontario, Canada

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 the construction of an online calculator.

Contributors 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 contextualised 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 contextualising 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, HJS 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 The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests 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.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement No additional data available.

REFERENCES

1. www.gradeworkinggroup.org [Accessed 27 Nov 2016].
2. Atkins D, Best D, Briss PA, *et al*. Grading quality of evidence and strength of recommendations. *BMJ* 2004;328:1490.
3. Balshem H, Helfand M, Schünemann HJ, *et al*. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64:401–6.
4. Hultcrantz M, Rind D, Akl EA, *et al*. The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol* 2017;87:4–13.
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.
6. Schunemann H, Brozek J, Guyatt G, *et al*. 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). Available: gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.s1fsemn2yjpm [Accessed 27 Nov 2016].
7. Guyatt GH, Oxman AD, Kunz R, *et al*. Going from evidence to recommendations. *BMJ* 2008;336:1049–51.
8. Schunemann H, Brozek J, Guyatt G, *et al*. 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. Available: http://

gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.i5hfweocv3qs [Accessed 27 Nov 2016].
9. Schunemann H, Brozek J, Guyatt G, *et al*. 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. Available: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.feb5g22tf6nv [Accessed 27 Nov 2016].
10. Schunemann H, Brozek J, Guyatt G, *et al*. 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). Available: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.y0ey8465h47v [Accessed 27 Nov 2016].
11. Schunemann H, Brozek J, Guyatt G, *et al*. 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. Available: http://gdt.guidelinedevelopment.org/app/handbook/handbook.html#h.xr5ac2p2khuq [Accessed 27 Nov 2016].
12. Sawaya GF, Guirguis-Blake J, LeFevre M, *et al*. 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.
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.
14. Schunemann H, Brozek J, Guyatt G, *et al*. 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. Available: https://gdt.gradepro.org/app/handbook/handbook.html#h.1i2bwkm8zpjo [Accessed 6 Apr 2019].
15. Hazlewood GS, Bombardier C, Tomlinson G, *et al*. 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.
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, *et al*. 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 edn. New York, NY: Chapman and Hall/CRC, 1993.
21. Spiegelhalter DJ, Myles JP, Jones DR, *et al*. An introduction to Bayesian methods in health technology assessment. *BMJ* 1999;319:508–12.
22. Zhang Y, Alonso-Coello P, Guyatt GH, *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 all do no harm!. *J Clin Pharmacol* 2005;45:371–7.
24. Schünemann HJ. Guidelines 2.0: do no net harm-the future of practice guideline development in asthma and other diseases. *Curr Allergy Asthma Rep* 2011;11:261–8.