BMJ Open

Defining decision thresholds for judgments on health benefits and harms using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Evidence to Decision (EtD) frameworks: a protocol for a randomised methodological study (GRADE-THRESHOLD)

Gian Paolo Morgano,[1,2] Lawrence Mbuagbaw ,[1] Nancy Santesso,[1,2] Feng Xie ,[1] Jan L Brozek,[1,2] Uwe Siebert,[3,4,5] Antonio Bognanni,[1] Wojtek Wiercioch,[1,2] Thomas Piggott,[1,2] Andrea J Darzi ,[1,2] Elie A Akl ,[6] Ilse M Verstijnen,[7] Elena Parmelli ,[8] Zuleika Saz-Parkinson,[8] Pablo Alonso-Coello ,[9] Holger J Schünemann[1,2,10,11]

To cite: Morgano GP, Mbuagbaw L, Santesso N, *et al*. Defining decision thresholds for judgments on health benefits and harms using the Grading of Recommendations Assessment, Development and Evaluation (GRADE) Evidence to Decision (EtD) frameworks: a protocol for a randomised methodological study (GRADE-THRESHOLD). *BMJ Open* 2022;**12**:e053246. doi:10.1136/bmjopen-2021-053246

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2021-053246).

Received 12 May 2021
Accepted 07 February 2022

Check for updates

© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to
Dr Holger J Schünemann; schuneh@mcmaster.ca

ABSTRACT

Introduction The Grading of Recommendations Assessment, Development and Evaluation (GRADE) and similar Evidence to Decision (EtD) frameworks require its users to judge how substantial the effects of interventions are on desirable and undesirable people-important health outcomes. However, decision thresholds (DTs) that could help understand the magnitude of intervention effects and serve as reference for interpretation of findings are not yet available.

The objective of this study is an approach to derive and use DTs for EtD judgments about the magnitude of health benefits and harms. We hypothesise that approximate DTs could have the ability to discriminate between the existing four categories of EtD judgments (Trivial, Small, Moderate, Large), support panels of decision-makers in their work, and promote consistency and transparency in judgments.

Methods and analysis We will conduct a methodological randomised controlled trial to collect the data that allow deriving the DTs. We will invite clinicians, epidemiologists, decision scientists, health research methodologists, experts in Health Technology Assessment (HTA), members of guideline development groups and the public to participate in the trial. Then, we will investigate the validity of our DTs by measuring the agreement between judgments that were made in the past by guideline panels and the judgments that our DTs approach would suggest if applied on the same guideline data.

Ethics and dissemination The Hamilton Integrated Research Ethics Board reviewed this study as a quality improvement study and determined that it requires no further consent. Survey participants will be required to read a consent statement in order to participate in this study at the beginning of the trial. This statement reads: You are being invited to participate in a research project which aims to identify indicative DTs that could assist users of the GRADE EtD frameworks in making judgments. Your input will be used in determining these indicative thresholds. By completing this survey, you provide consent that the anonymised data collected will be used for the research study and to be summarised in aggregate in publication and electronic tools.

PROTOCOL registration number NCT05237635.

Strengths and limitations of this study

- The calculation of the decision thresholds will be based on empirical data.
- We will use structured case scenarios to present survey participants with the information relevant to make their judgments.
- We will employ a randomisation process to ensure that case scenarios will be equally distributed across survey participants.
- We acknowledge that the survey requires effort and that this could impact test–retest reliability and applicability of the survey results which we overcome in part by conducting a large trial.

INTRODUCTION

As advocated by the National Academy of Medicine of the United States (formerly the Institute of Medicine), the assessment of the benefits and harms of alternative care options

BMJ Open: first published as 10.1136/bmjopen-2021-053246 on 10 March 2022. Downloaded from http://bmjopen.bmj.com/ on April 14, 2024 at American University of Beirut. Protected by copyright.

(ie, interventions, actions) is an essential component of any decision-making process underlying guideline recommendations.[1] This assessment should be explicit and include considerations around the probability, magnitude and importance of health-related benefits and health-related harms, and other desirable and undesirable consequences of the recommendation or decision.[2] The Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group has developed the Evidence to Decision (EtD) frameworks to help guideline developers use the evidence in a structured and transparent way and to ensure that they consider all the criteria relevant to their decisions.[3 4] The GRADE and other EtD frameworks require decision-makers to evaluate explicitly the benefits and harms of alternative care options through separate judgments based on the two following questions: '*How substantial are the desirable anticipated effects (health benefits)?*', '*How substantial are the undesirable anticipated effects (health harms)?*'. The guidance from the GRADE Working Group includes expressing and facilitating these judgments by assigning the health benefits or health harms of some intervention under evaluation to one of the following four categories: 'Trivial or None', 'Small', 'Moderate' and 'Large'.[3 4] To be useful, however, this simplification requires that EtD users have a similar understanding of what magnitude of health benefits or health harms belong into which category and are consistent in their judgments. A similar common understanding is also important between those assigning a category and those interpreting the meaning of a category that is communicated to them (ie, 'imagining' how substantial is an effect based on the category). This can be achieved only when people make similar judgments. To direct EtD users on how to make these judgments appropriately, the GRADE Working Group has produced guidance articles that include the description of the underpinning concepts and examples of judgments based on clinical scenarios.[4 5] Despite the popular use of thresholds to support decision-making in various fields of healthcare research,[6–8] and its adoption by the GRADE approach,[9 10] use of decision thresholds (DTs) for EtD judgments about health benefits and harms is not yet established. For continuous outcomes, EtD users can revert to statistical notions such as Cohen's standardised effect sizes or the minimal important difference to interpret the magnitude of effects.[11 12] However, empirical data supporting judgments on health benefits and harms for dichotomous outcomes are not yet available for the EtDs.

Objectives

The objective of this study is to derive DTs for EtD judgments on the magnitude of health benefits and harms. We hypothesise that DTs could discriminate between the four categories for EtD judgments. Explicit DTs, providing an indication for which could be the appropriate judgement for a given scenario, might have the potential to support panels of decision-makers in their work, facilitate a common understanding and promote consistency and transparency in judgments.

Conceptual approach

In the proposed DTs approach, we will consider that judgments on how substantial anticipated effects (health benefits and harms) are should be influenced by: (1) the size of the intervention's effects on each outcome (eg, the probability of people who experience benefit or harm); and (2) the value assigned to those outcomes by the people who are affected.[5] Under this assumption, we will collect data about the association between the dyad composed of size of intervention's effects and value of the outcome on one hand, and judgments on the magnitude of the anticipated effects on the other. In accordance with the EtD frameworks, judgments on desirable effects and on undesirable effects will be collected separately and should not account for any potential tradeoff between benefits and harms. We will use this data to estimate the DTs and provide a conceptual framework for their interpretation and use (see online supplemental file 1).

METHODS

This study will consist of two parts. In the first part, we will conduct a methodological randomised controlled trial to collect the data that will be used to derive the DTs. Second, we will investigate the validity of our DTs by measuring the agreement between judgments that were made in the past by guideline panels and the judgments that our DTs approach would suggest if applied on the same guideline data.

Randomizsed controlled trial

The following description of methods and analysis of this trial follows the latest guidance by the Standard Protocol Items: Recommendations for Interventional Trials.[13] We registered this protocol in the Protocols Registration and Results System (clinicaltrials.gov, protocol registration number: NCT05237635).

Design and setting

Study participants will be recruited to complete a randomised electronic survey (see online supplemental file 2) designed to elicit ratings on the magnitude of the potential health effects (benefits or harms) of interventions. Ratings on health benefits and health harms will be collected separately. We will organise the survey into three sections: introduction and example, ratings and questions about respondent demographics. Ratings will be based on five outcomes having a different impact on health (death, major ischaemic stroke, pulmonary embolism of moderate severity, diarrhoea of moderate severity and mild nausea/vomiting) presented through descriptive case scenarios. Each case scenario will include: (1) a GRADE Summary of Finding table[14] providing

BMJ Open: first published as 10.1136/bmjopen-2021-053246 on 10 March 2022. Downloaded from http://bmjopen.bmj.com/ on April 14, 2024 at American University of Beirut. Protected by copyright.

information about the PICO (Population, Intervention, Comparator, Outcome), the relative and absolute anticipated effects of the intervention and the certainty in the evidence; (2) a Health Outcome Descriptor[15] describing key attributes of the outcome under consideration including symptoms, time horizon, testing and treatment and consequences; and (3) a measure of the impact on health of the outcome (also known as 'value' of the outcome or 'health utility' in health economics). This measure will be expressed on a scale from 0 (being dead) to 1 (perfect health) which means that outcomes with a higher value are valued closer to perfect health as compared with outcomes with a lower value. For each outcome, we will include a case scenario descriptive of desirable health effects and another one descriptive of undesirable health effects, for a total of 10 case scenarios across five outcomes. These scenarios differ in the description of the severity of the outcome and the consequences to represent clearly different values.

Participants

Selection criteria

The target population of the survey will include clinicians, epidemiologists, decision scientists, health research methodologists, experts in HTA and members of guideline working groups, but it will be open to the public too. Prior knowledge of the GRADE approach and experience with the EtD frameworks will not be required for participation.

Patient and public involvement statement

There was no direct dedicated patient or public involvement but patients and the public will participate in the survey and can provide feedback.

Recruitment

We will distribute the survey through colleagues, the research group's e-mail lists including that of the Cochrane Collaboration, Guidelines International Network, guideline developers and of the Global Evidence Synthesis Initiative. Twitter, LinkedIn and other social medial platforms will also be used for broader distribution. We will continue recruitment for this trial until reaching our anticipated sample size (see below) or until 31 December 2022 as it is unlikely that we will meet the sample size through additional recruitment efforts beyond then.

Intervention and comparison

Participants will be randomised to a set of four case scenarios, written in lay language, that will be used as intervention (or comparison) in this trial. For each case scenario, we will ask survey participants to consider the intervention's effects and the value of the outcome and rate how substantial the described health benefits or health harms are. We will also ask them to indicate the lower and upper bound for the ranges of magnitudes of absolute risk difference (ARD) that they associate with the judgments of 'Small' and 'Moderate'. Any estimate below the lower bound for 'Small' will be considered as 'Trivial or None', and any estimate above the upper bound of 'Moderate' will be considered as 'Large'.

Outcomes

The primary endpoints of this trial are the three DTs (T1=DTTrivial/Small, T2=DTSmall/Moderate, T3=DT-Moderate/Large) that would allow discriminating between EtD judgments of 'Trivial or None' and 'Small', 'Small' and 'Moderate' and 'Moderate' and Large', respectively.

Randomizsation

Randomisation will ensure that case scenarios will be equally distributed across survey participants to get balanced judgments on outcomes. It will reduce potential confounding due to order effects and possible differences between case scenarios (eg, clarity). Randomisation will also avoid selection bias that could arise if allowing participants to select the case scenarios more familiar to them.

Sample size calculation

We based our sample size calculation on the data collected during pilot testing (n=15 participants). Based on this data, we computed the mean thresholds T1, T2 and T3 for each outcome separately and estimated that we need to recruit 1406 survey respondents to demonstrate a difference of 15% of the mean with non-overlapping 95% CIs. These computations were done using Winpepi.[16]

Statistical methods

Calculation of thresholds from survey ratings

We will use the ranges of ARD for judgments of 'Small' and 'Moderate' collected through the survey to calculate the thresholds associated with each rating. The thresholds will be derived through the product between each ARD indicated as range boundary and the difference in value from perfect health (1 — outcome's value) for the outcome associated with that rating (see online supplemental file 2). We will calculate the DTs as the weighted mean of the corresponding thresholds derived from survey ratings. We will use a weighted mean to account for multiple ratings from the same survey respondent.

Primary analysis

We will use frequencies and percentages to describe the characteristics of survey respondents. For each DT, we will calculate mean, SD and 95% CIs. We will conduct an analysis of variance (ANOVA) to determine if there are any differences between the thresholds (T1≠T2≠T3). If we identify a difference, since each participant will contribute data to each threshold, we will employ a post-hoc paired sample t-test to assess which of the DTs are different that is, (T1≠T2; T2≠T3; T1≠T3). Our a-priori hypothesis is that there will

BMJ Open: first published as 10.1136/bmjopen-2021-053246 on 10 March 2022. Downloaded from http://bmjopen.bmj.com/ on April 14, 2024 at American University of Beirut. Protected by copyright.

be a difference between the DTs and no difference between the magnitude of DTs for benefits and harms.

Sensitivity analyses

Subgroup analysis

We will conduct explorative subgroup analyses based on participants' characteristics (training in epidemiology, familiarity with the EtD frameworks, previous participation in guideline development groups, language used). Our a-priori hypotheses is that, in each of the identified subgroups, there will be a difference between the DTs and no difference between the magnitude of DTs for benefits and harms.

Incoherent ratings and outliers

We expected that, given the complexity of the topic, some responses might not be internally coherent or represent outliers. We define a threshold as incoherent if T1>T2 or T2>T3. We define thresholds as outliers if they fall more than three IQRs below the first quartile or above the third quartile. We will verify if the primary analysis would differ if incoherent thresholds or data outliers are excluded. The a-priori hypothesis for the sensitivity analyses will be the same as for the primary analysis.

Order effects

We will conduct an ANOVA analysis to assess for potential order effects. We will examine whether participants randomised to a case scenario for a low-value outcome (outcome value <0.5) in the first case-scenario provided different thresholds as compared with participants who were randomised to a high-value outcome first. Similarly, we will examine whether participants who provided a judgement of 'Small' in the first iteration provided different thresholds as compared with participants who provided a judgement of 'Large' in the first iteration. Our a-priori hypothesis is that of no differences if comparing each DT between these groups.

Retrospective comparison of judgments

To investigate the validity of our DTs, we will purposively select judgments from existing guidelines developed using the EtD frameworks and measure the agreement between judgments made by guideline panels and the judgments that our DTs approach would suggest. We will consider for inclusion guidelines reporting the value assigned to outcomes during the decision-making process. We will use frequencies and percentages to describe the agreement. We will employ SPSS V.26 (IBM Corp, Armonk, New York, USA) to conduct all statistical analyses. We will use the Bonferroni correction for multiple testing in all secondary analyses.[17]

Pilot testing and assessment of feasibility

To ensure usability and clarity of the survey across respondents having different background or expertise, we piloted the survey with study co-investigators as well as complementary representatives of the target population (n=15). Comments on three iterations of the survey were collected either electronically or by voice recordings and discussed during study meetings. Furthermore, to test the feasibility of the study, we recruited 75 participants from the target population. Participants were able to complete the exercise in the majority of cases. Only 7 out of 75 did not complete the survey after they signed up. Participants contributed a total of 295 ratings with only 17 out of 312 expected ratings missing indicating that the approach to obtaining DTs is feasible. This is true for people of varying backgrounds and educational levels. The findings based on the preliminary analysis of the data support our hypothesis that DTs can help discriminate between the judgments (see online supplemental file 3). Furthermore, we will use periodic interim results to inform judgments by guideline groups that develop recommendations but will not use these to draw final conclusions about the trial results until it is stopped formally by reaching the calculated sample size or on 31 December 2022. No additional data are available.

ETHICS AND DISSEMINATION

After review, the Hamilton Integrated Research Ethics Board (HiREB) determined that as a quality improvement project, this study was exempt from formal ethics review as per TCPS2 (2014) Article 2.5. We will inform respondents of this decision and the anonymous nature of the study. Survey participants will be required to read a consent statement in order to participate in this study at the beginning of the trial. This statement reads: You are being invited to participate in a research project which aims to identify indicative DTs that could assist users of the GRADE EtD frameworks in making judgments. Your input will be used in determining these indicative thresholds. By completing this survey, you provide consent that the anonymised data collected will be used for the research study and to be summarised in aggregate in publication and electronic tools.

The results of this randomized trial will be published in a peer-reviewed journal. We also aim to present the results in national and international conferences.

DISCUSSION

We believe that DTs for judgments on desirable and undesirable health effects can be useful to decision-makers using the EtD frameworks. Guideline panels using the GRADE EtDs often ask what are 'Trivial or None', 'Small', 'Moderate' and 'Large' effects. The proposed DTs approach could provide an answer based on empirical data and be used to initiate and promote discussion. Furthermore, it is simple to apply, and requires only to calculate the product between ARD and the reduction in value associated with the outcome. This endeavour will expand the research on the use of DTs within the GRADE methodology and could be integrated into GRADEpro.

BMJ Open: first published as 10.1136/bmjopen-2021-053246 on 10 March 2022. Downloaded from http://bmjopen.bmj.com/ on April 14, 2024 at American University of Beirut. Protected by copyright.

Our work with Hultcrantz *et al*[10] suggests that clinical DTs can be used to allow appropriate ratings of the certainty of the evidence, but there is no empirical data. Furthermore, it focuses on the construct of certainty of evidence and targets different degrees of contextualisation, while we address judgments on the magnitude of effects and made by users of the EtD frameworks. The joint consideration of the estimate of effect and outcome's importance has been already adopted in another effort of the GRADE Working Group. In a GRADE concept paper,[18] Alper *et al* aim to define the certainty in the net benefit and suggest calculating the net effect of an intervention by combining importance-adjusted effect estimates calculated from different outcomes. While this strategy is appealing and would allow us to apply our research to EtD judgments on the trade-off between benefits and harms, further research is needed to establish if the estimates to be combined are independent and not correlated with each other. Other quantitative approaches to assess the benefits, harms and net benefit associated with treatments are available in the literature,[19] but none aims to characterise the magnitude of effects into categories (ie, 'Trivial or None', 'Small', 'Moderate', 'Large') as needed to make judgments using the EtD frameworks. Utilitarian frameworks are common in health economic research, where health-utilities elicited from target populations are used to inform modelling techniques such as cost-effectiveness analysis based on quality-adjusted-life-years.[20 21] However, our trial will not be free of limitations. Generalisability of the findings may be limited by the use of the case scenarios we chose and the limited number of effect sizes we include in the trial. Generalisability may also be limited by the type of participants we will be able to recruit. Therefore, we plan, following the completion of this trial, to conduct further research with additional case scenarios and different target populations.

Author affiliations
[1]Department of Health Research Methods, Evidence and Impact, McMaster University, Hamilton, Ontario, Canada
[2]Michael G. DeGroote Cochrane Canada & McMaster GRADE Centres, McMaster University, Hamilton, Ontario, Canada
[3]Institute of Public Health, Medical Decision Making and Health Technology Assessment, Department of Public Health, Health Services Research and Technology Assessment, UMIT – University for Health Sciences, Medical Informatics and Technology, Hall in Tirol, Austria
[4]Division of Health Technology Assessment and Bioinformatics, ONCOTYROL - Center for Personalized Cancer Medicine, Innsbruck, Austria
[5]Center for Health Decision Science/Department of Health Policy and Management, Harvard TH Chan School of Public Health and MGH-ITA/Department of Radiology, Harvard Medical School, Harvard University, Boston, Massachusetts, USA
[6]Department of Internal Medicine, American University of Beirut, Beirut, Lebanon
[7]Dutch National Health Care Institute, Diemen, The Netherlands
[8]European Commission Joint Research Centre, Ispra, Italy
[9]Iberoamerican Cochrane Centre - Department of Clinical Epidemiology and Public Health, Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain
[10]Department of Medicine, McMaster University, Hamilton, Ontario, Canada
[11]Dipartimento di Scienze Biomediche, Humanitas University, Milan, Italy

Twitter Feng Xie @fengxie_mac

Contributors HJS is the principal investigator and conceived of the study and together with GPM, NS, FX and JLB designed and established this research project. LM, IMV, TP, EP, ZS-P, AB and US piloted the survey and provided methodological input. GPM was responsible for the ethics application and with HJS for registration of the protocol on clinicaltrials.gov. HJS, GPM, AB and LM designed the statistical analysis. GPM, AB, WW, AJD and HJS are responsible for recruitment. HJS, WW and GPM are responsible for the coordination of the study. GPM and HJS drafted the manuscript. NS, FX, AB and JLB reviewed early drafts. EAA and PA-C provided methodological input in the study design. All listed authors participated in the writing and revision of the manuscript and approved its final version.

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 HJS is the co-chair of the GRADE working group. Decision thresholds will be used in the GRADEpro app and for other projects. Currently no financial interests. HJS and JLB are codevelopers of the GRADEpro app. US is an unpaid member of Working Group for the German Clinical S3 Guideline Prevention of Cervical Cancer; Committee for Cancer Screening of the Austrian Federal Ministry of Health; Oncology Advisory Council of the Federal Ministry of Health, Austria.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Lawrence Mbuagbaw http://orcid.org/0000-0001-5855-5461
Feng Xie http://orcid.org/0000-0003-3454-6266
Andrea J Darzi http://orcid.org/0000-0002-2498-1697
Elie A Akl http://orcid.org/0000-0002-3444-8618
Elena Parmelli http://orcid.org/0000-0002-3730-8530
Pablo Alonso-Coello http://orcid.org/0000-0002-8001-8504

REFERENCES

1 Institute of Medicine Committee on Standards for Developing Trustworthy Clinical Practice G. *Clinical practice guidelines we can trust*. Washington (DC): National Academies Press (US), 2011.
2 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.
3 Atkins D, Eccles M, Flottorp S, *et al*. Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches the grade Working group. *BMC Health Serv Res* 2004;4:38.
4 Alonso-Coello P, Schünemann HJ, Moberg J, *et al*. Grade evidence to decision (ETD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: introduction. *BMJ* 2016;353:i2016.
5 Alonso-Coello P, Oxman AD, Moberg J, *et al*. Grade evidence to decision (ETD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: clinical practice guidelines. *BMJ* 2016;353:i2089.
6 Marcucci M, Sinclair JC. A generalised model for individualising a treatment recommendation based on group-level evidence from randomised clinical trials. *BMJ Open* 2013;3:e003143.
7 McCabe C, Claxton K, Culyer AJ. The NICE cost-effectiveness threshold: what it is and what that means. *Pharmacoeconomics* 2008;26:733–44.
8 Pauker SG, Kassirer JP. Therapeutic decision making: a cost-benefit analysis. *N Engl J Med* 1975;293:229–34.

BMJ Open: first published as 10.1136/bmjopen-2021-053246 on 10 March 2022. Downloaded from http://bmjopen.bmj.com/ on April 14, 2024 at American University of Beirut. Protected by copyright.

9 Hultcrantz M, Mustafa RA, Leeflang MMG, *et al*. Defining ranges for certainty ratings of diagnostic accuracy: a grade concept paper. *J Clin Epidemiol* 2020;117:138–48.
10 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.
11 Faraone SV. Interpreting estimates of treatment effects: implications for managed care. *P T* 2008;33:700–11.
12 Johnston BC, Ebrahim S, Carrasco-Labra A, *et al*. Minimally important difference estimates and methods: a protocol. *BMJ Open* 2015;5:e007953.
13 Chan A-W, Tetzlaff JM, Gøtzsche PC, *et al*. Spirit 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ* 2013;346:e7586.
14 Guyatt GH, Thorlund K, Oxman AD, *et al*. Grade guidelines: 13. preparing summary of findings tables and evidence profiles-continuous outcomes. *J Clin Epidemiol* 2013;66:173–83.
15 Baldeh T, Saz-Parkinson Z, Muti P, *et al*. Development and use of health outcome descriptors: a Guideline development case study. *Health Qual Life Outcomes* 2020;18:167.
16 Abramson JH. WINPEPI updated: computer programs for epidemiologists, and their teaching potential. *Epidemiol Perspect Innov* 2011;8:1.
17 Abdi H. Holm's sequential Bonferroni procedure. In: Salkin N, ed. *Encyclopedia of research design*. Thousand Oaks, California, 2010: 1–8.
18 Alper BS, Oettgen P, Kunnamo I, *et al*. Defining certainty of net benefit: a grade concept paper. *BMJ Open* 2019;9:e027445.
19 Puhan MA, Singh S, Weiss CO, *et al*. A framework for organizing and selecting quantitative approaches for benefit-harm assessment. *BMC Med Res Methodol* 2012;12:173.
20 Siebert U. When should decision-analytic modeling be used in the economic evaluation of health care? *Eur J Health Econ* 2003;4:143–50.
21 Weinstein MC, Stason WB. Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med* 1977;296:716–21.

BMJ Open: first published as 10.1136/bmjopen-2021-053246 on 10 March 2022. Downloaded from http://bmjopen.bmj.com/ on April 14, 2024 at American University of Beirut. Protected by copyright.