Check for updates

Core GRADE 2: choosing the target of certainty rating and assessing imprecision

Gordon Guyatt,[1,2,3,*] Linan Zeng,[3,4,5,6,*] Romina Brignardello-Petersen,[1] Manya Prasad,[7] Hans De Beer,[8] M Hassan Murad,[9] Alfonso Iorio,[1,2] Arnav Agarwal,[1,2,3] Liang Yao,[10] Thomas Agoritsas,[1,3,11] Jamie Rylance,[12] Reem A Mustafa,[1,13] Per Olav Vandvik,[3,14] Prashanti Eachempati,[3,15,16] Chunjuan Zhai,[17] Lingli Zhang,[4,5,6] Victor M Montori,[18,19] Monica Hultcrantz[20,21]

For numbered affiliations see end of the article

Correspondence to: G Guyatt guyatt@mcmaster.ca; (ORCID 0000-0003-2352-5718)

Additional material is published online only. To view please visit the journal online.

Cite this as: *BMJ* 2025;389:e081904
http://dx.doi.org/10.1136/bmj-2024-081904

Accepted: 14 March 2025

This second article in a seven part series presents the Core GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach to deciding on the target of the certainty rating, and decisions about rating down certainty of evidence due to imprecision. Core GRADE users assess if the true underlying treatment effect is important or not in relation to the minimal important difference (MID) or, alternatively, if a true underlying treatment effect exists. The location of the point estimate of effect in relation to the chosen threshold determines the target. For instance, using the MID thresholds, a point estimate greater than the MID suggests an important effect and less than the MID, an unimportant or little to no effect. Users then rate down for imprecision if the 95% confidence interval crosses the MID for benefit or harm.

SUMMARY POINTS

Core GRADE (Grading of Recommendations Assessment, Development and Evaluation) users may assess their certainty in an important or unimportant effect (using the minimal important difference threshold), which often simplifies and clarifies ratings of imprecision and interpretation of results

In some situations, when wanting to minimise value judgments, Core GRADE users may use the null threshold to assess their certainty in a true underlying treatment effect

When deciding whether to rate down certainty for imprecision, Core GRADE users will consider whether the confidence interval (CI) crosses the chosen threshold and will usually rate down two levels if the CI includes both an important benefit and an important harm

If the CI does not cross a threshold but the effect is very large, Core GRADE users will consider rating down for imprecision if the sample size or number of events are too small (ie, not meeting the optimal information size)

This is the second paper in a series describing the essentials of GRADE (Grading of Recommendations Assessment, Development and Evaluation) in rating certainty of evidence and grading recommendations, focusing on evidence addressing alternative care options from the perspective of the patient. In the first paper, we described our overall Core GRADE approach, noting that Core GRADE's certainty rating represents confidence that the true effect lies on one side of a threshold (such as an important difference) or in a particular range (such as an unimportant or little to no effect).[1] We further noted that randomised controlled trials start as high certainty evidence in Core GRADE's four category certainty approach (high, moderate, low, and very low), and non-randomised studies of interventions (observational) start at low certainty.

This paper comprises two parts. The first part describes a key element of the Core GRADE approach: choosing the target of the certainty rating—that is, what it is in which we are rating our certainty.[2 3] The second part provides guidance on decisions about rating down for imprecision—one of the five domains that Core GRADE users consider when assessing the certainty of evidence in both randomised controlled trials and non-randomised studies of interventions.[4] The following papers in this series deal with the other four domains.

The information in this article will enable Core GRADE users to make choices about whether to assess their certainty in an important or unimportant effect, or assess their certainty in a true underlying treatment effect; make judgments about rating down certainty of evidence once or twice for imprecision by considering the width the confidence interval (CI) and the appropriate plain language summary; and decide when and how to consider the optimal information size in making judgments about rating down certainty of evidence for imprecision.

Target of the certainty rating

When assessing the effect of an intervention, the primary interest is whether it outperforms alternatives such as standard care or other existing treatments. If no difference exists in benefit outcomes, a guideline panel will unlikely recommend and clinicians will unlikely use the new treatment unless it offers other advantages, such as reduced harms or burdens. Moreover, merely identifying the presence of an effect is often insufficient to recommend a treatment: patients and clinicians need to know whether the effect is large

BMJ: first published as 10.1136/bmj-2024-081904 on 29 April 2025. Downloaded from https://www.bmj.com/ on 16 August 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

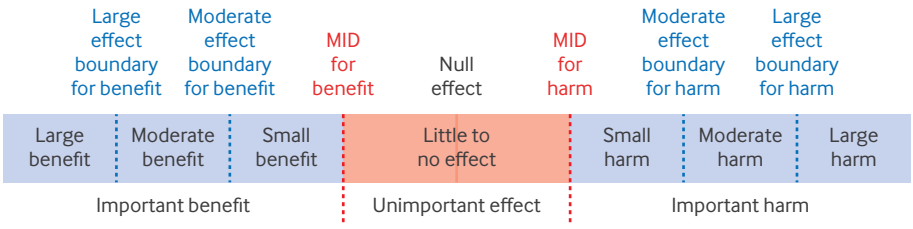

Fig 1 | Thresholds and ranges for rating certainty of evidence in Core GRADE. Besides the Core GRADE thresholds of null effect and MID, two other thresholds may be considered—the moderate effect threshold that demarcates small versus moderate effects, and the large effect threshold that demarcates moderate versus large effects. GRADE=Grading of Recommendations Assessment, Development and Evaluation; MID=minimal important difference

enough to be important. The question of whether there is an effect compared with the alternative corresponds to using the threshold of null effect, whereas the question of whether the effect is important aligns with using the minimal important difference (MID). The MID, a crucial concept in clinical studies and Core GRADE methodology, represents the smallest change in a single outcome that patients perceive as important.[5]

The focus of Core GRADE is on these two questions: whether there is an effect compared with the alternative (ie, using the null threshold) and whether the effect is large enough to be important for patients (ie, using the MID) (fig 1). Use of additional thresholds of moderate and large effects has proved challenging for GRADE users and in our judgment does not provide important incremental value in making sound and optimally useful ratings of certainty. Nevertheless, supplementary appendix 1 presents judgments of small, moderate, and large effects that Core GRADE users may sometimes want to consider but are not part of Core GRADE.[6]

Deciding what it is in which we are rating our certainty requires three steps (fig 2). For the first step, Core GRADE users choose if they are interested in whether an effect is or is not important, or whether a true underlying effect compared with the alternative exists.

For the next step, Core GRADE users establish the effect estimates through meta-analysis. An important choice they face is whether to use fixed effect or random effect statistical models in their analysis. Supplementary appendix 2 provides key issues with the approaches.

In rating certainty of evidence, Core GRADE users typically consider absolute rather than relative effects.

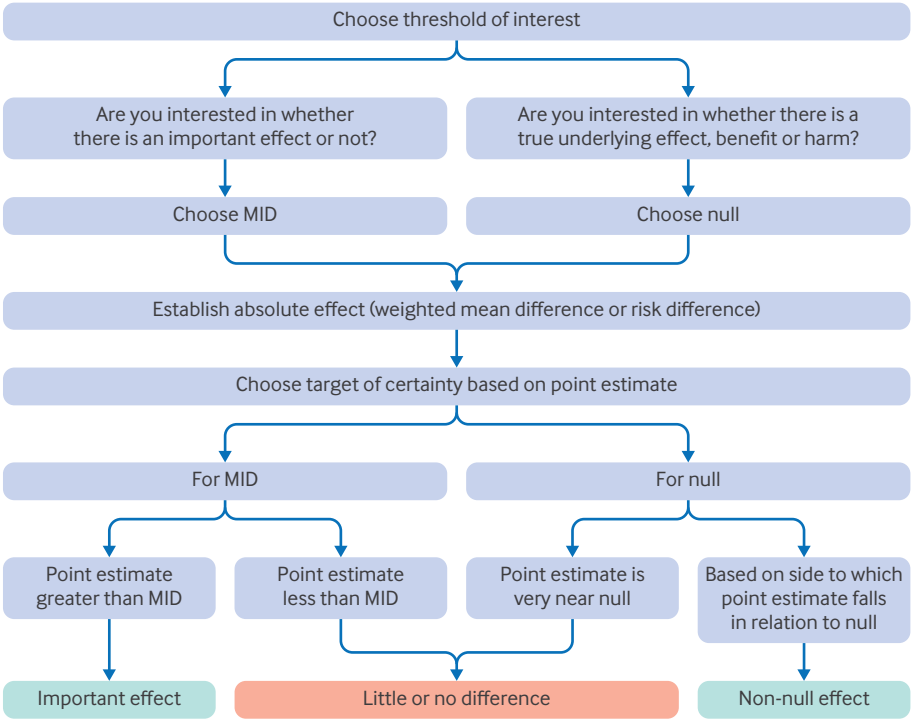

Fig 2 | Core GRADE steps for deciding target of certainty rating. GRADE=Grading of Recommendations Assessment, Development and Evaluation; MID=minimal important difference

BMJ: first published as 10.1136/bmj-2024-081904 on 29 April 2025. Downloaded from https://www.bmj.com/ on 16 August 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

For binary outcomes, Core GRADE users obtain the best estimate of the risk difference and its 95% CI by applying the relative risk to an estimate of the baseline risk. Supplementary appendix 3 illustrates how this is done. A subsequent paper in this series focusing on summary of findings tables presents additional information about absolute effects and how to present continuous outcomes.[7]

Finally, Core GRADE users assess the magnitude of the absolute effect estimate in relation to the chosen threshold. The process for these steps is discussed further below and illustrated in figure 2.

Assessing if the effect is or is not important (using MID threshold)

When considering whether an effect is important, Core GRADE users must focus on absolute (ie, risk differences) rather than relative effects. The reason is, as we explained in the first article in this series, it is absolute rather than relative effects that are important to patients—a 50% relative risk reduction (risk ratio of 0.5) could represent a 1% absolute reduction (from a baseline risk of 2% in control group to 1% in intervention group) or a much larger 20% absolute reduction (from 40% to 20%).

If Core GRADE users are interested in whether an effect is important, they will thus need to make a value judgment about the importance of the outcome, and, in particular, the threshold that delineates an important from an unimportant effect (ie, the MID). The values and preferences that drive this choice should be those of the patients or other target populations, such as the general public.

Guideline development and health technology assessment require judgments about how people value the benefits, harms, and burdens of the interventions under consideration. Specifying MIDs, using either established MIDs (most likely to be available for patient reported outcomes such as pain, functional status, or quality of life[5 8]) or their own estimates (generated from, for example, existing literature or their clinical experience[9]) has proved helpful in facilitating the trade-offs between desirable and undesirable consequences of interventions. In the seventh and final article in this series we address in some detail the issues of choosing MIDs for key outcomes.[10]

If Core GRADE users have chosen the MID threshold, and the point estimate from the meta-analysis represents an effect greater than the MID ((a) in figure 3), systematic review authors will rate their certainty that the true effect is an important benefit (or, if favouring the comparator, an important harm). If the point estimate represents an effect less than the MID ((b) in figure 3), they will rate their certainty in an unimportant (little to no) effect.[3]

Assessing whether a true underlying treatment effect exists (using null as threshold)

For several reasons, the null represents an alternative threshold to the MID in systematic reviews. Evidence on the distribution of values and preferences in the population of interest is typically limited, making inferences about the MID challenging. Furthermore, systematic review authors may not see their mandate as including the search and interpretation of relevant evidence about MIDs. Finally, systematic review authors may want to leave the value judgments involved in choosing specific MIDs to health technology assessment and guideline practitioners who typically consult a wider group of individuals, and often in a structured way.

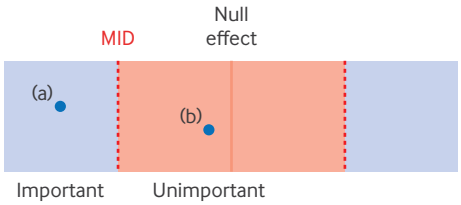

Fig 3 | Using Core GRADE to assess if an effect is or is not important. (a) As the point estimate is above the MID, the target of the certainty rating is that the true effect is important. (b) As the point estimate is below the MID, the target of the certainty rating is that the true effect is unimportant (little to no effect). GRADE=Grading of Recommendations Assessment, Development and Evaluation; MID=minimal important difference

If Core GRADE users have chosen the null they will, based on where the point estimate falls in relation to the null, typically rate certainty that a true beneficial or a harmful effect exists. If, however, the point estimate is near the null, because the intuitive inference in such situations is that the true effect represents little to no difference between intervention and control, they will rate their certainty in an unimportant effect ((b) in figure 3).[11] So, although choosing the null usually avoids specifying MIDs, it will not always do so. Supplementary appendix 4 provides an expanded discussion of this issue. The difficulty in altogether escaping considerations of importance when choosing the null as a threshold may lead Core GRADE users to prefer the MID as a threshold.

Rating certainty of evidence: imprecision

After deciding on the target of certainty rating, Core GRADE users assess whether limitations exist in any one of five GRADE domains (imprecision, inconsistency, risk of bias, indirectness, and publication bias). The following discussion addresses how Core GRADE users can make judgments about imprecision.

Imprecision defined

Studies of interventions seek to estimate the true underlying treatment effect. A meta-analysis provides our best estimate of the effect (the point estimate), and the CIs provide the bounds within which the true effect plausibly lies. The most commonly used CI is the 95% CI. The CI's width provides key information about the extent of imprecision, thus informing the impact of random error on certainty of evidence.

BMJ: first published as 10.1136/bmj-2024-081904 on 29 April 2025. Downloaded from https://www.bmj.com/ on 16 August 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

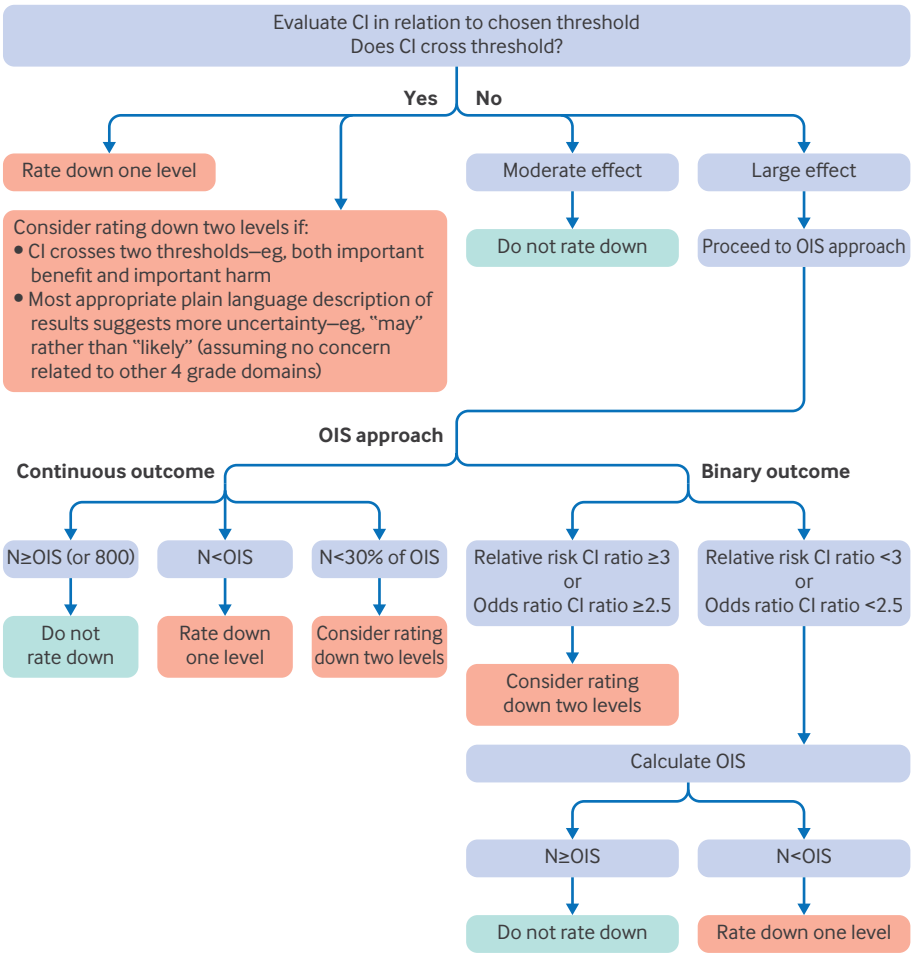

Fig 4 | Core GRADE steps for rating imprecision. The relative risk CI ratio represents the upper boundary divided by lower boundary of CI of relative risk, and the odds ratio CI ratio represents the upper boundary divided by lower boundary of CI of odds ratio. CI=confidence interval; GRADE=Grading of Recommendations Assessment, Development and Evaluation; N=number of participants; OIS=optimal information size

We will now describe our approach to making the judgment of whether the CI is sufficiently wide that Core GRADE users should rate down for imprecision, and whether they should rate down by one or two levels. As judgments about rating down once or twice have proved challenging for GRADE users, in the current discussion we build on clarification in previous papers.[4 6] Figure 4 presents the steps Core GRADE users take in making these judgments.

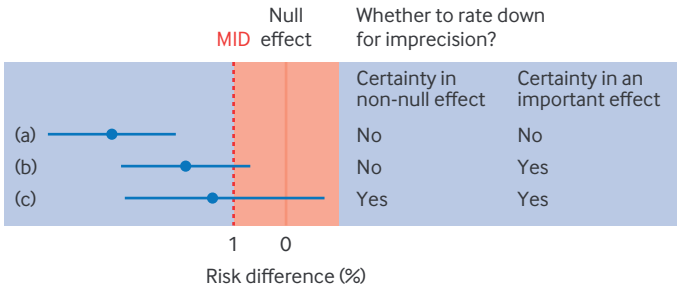

Fig 5 | Example of how the target of certainty rating using Core GRADE (above the MID or above the null) affects the rating of imprecision. (a) Core GRADE users will not rate down for imprecision in either case. (b) If the target of certainty rating is an important effect (above the MID), Core GRADE users will rate down for imprecision, but they will not if the target of certainty rating is an effect above the null. (c) Core GRADE users will rate down for imprecision in both cases. GRADE=Grading of Recommendations Assessment, Development and Evaluation; MID=minimal important difference

Rating down (or not) for imprecision

When deciding whether to rate down certainty for imprecision, Core GRADE users will consider whether the CI crosses the chosen threshold. For instance, consider the pooled effect estimate from a hypothetical systematic review of randomised controlled trials illustrated in figure 5. For (a) in figure 5, whether Core GRADE users are rating certainty for a non-null effect (null being a risk difference of 0%) or an important effect (the MID threshold of 1%), the CI does not cross either threshold and they will not rate down their certainty for imprecision. Assuming they have no concerns about the other four GRADE domains, they will have high certainty of a non-null effect as well as an important effect.

For (b) in figure 5, decisions about rating down certainty will differ depending on the threshold. When using the null, as the CI does not cross the threshold,

BMJ: first published as 10.1136/bmj-2024-081904 on 29 April 2025. Downloaded from https://www.bmj.com/ on 16 August 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

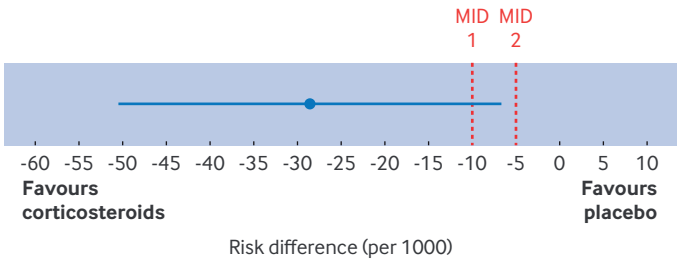

Fig 6 | An example of how rating down for imprecision in Core GRADE depends on the choice of MID in a systematic review of corticosteroids versus no corticosteroids on mortality in patients with community acquired pneumonia. If the review authors set the MID at MID1, they will rate down for imprecision, and if they set the MID at MID2, they will not rate down for imprecision. GRADE=Grading of Recommendations Assessment, Development and Evaluation; MID=minimal important difference

Core GRADE users will not rate down their certainty for imprecision. When using the MID, as the CI crosses the threshold, they will rate down for imprecision.

For (c) in figure 5, whether Core GRADE users are rating their certainty in relation to the null or the MID, the CI crosses the threshold and they will rate down for imprecision.

Consider a systematic review of corticosteroids versus no corticosteroids for patients with community acquired pneumonia (fig 6).[12] The meta-analysis of randomised controlled trials reported that corticosteroids yielded 29 fewer deaths per 1000 patients, with a CI from 52 fewer to 6 fewer. If review authors have chosen the null as their threshold, they will rate their certainty that a true mortality reduction exists and will not rate down for imprecision.

If review authors have chosen the MID as their threshold and set the MID at a difference of 10 deaths per 1000 patients (MID1 in figure 6), because the point estimate is greater than the threshold, they will rate down their certainty in an important mortality reduction. Had they chosen an MID of 5 deaths per 1000 patients (MID2 in figure 6), they would not rate down for imprecision because the CI does not cross the MID threshold.

When Core GRADE users have chosen the MID as their threshold and the point estimate is less than the MID, they will rate their certainty that the true treatment effect is unimportant (ie, little to no effect)

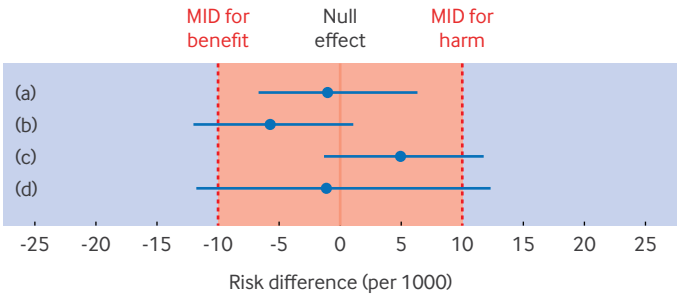

Fig 7 | Rating certainty in little to no effect and rating down for imprecision in Core GRADE when the confidence interval crosses the MID. GRADE=Grading of Recommendations Assessment, Development and Evaluation; MID=minimal important difference

(all point estimates in figure 7). As described in the section on assessing whether there is a true underlying treatment effect, when Core GRADE users have chosen the null as the threshold and the point estimate clearly suggests an unimportant effect (ie, the point estimate is close to the null) they will instead rate certainty in little to no effect by relating the CI to the MID (see supplementary appendix 4 for further illustration). In either case, they will not rate down for imprecision if the CI crosses neither threshold ((a) in figure 7). If the CI crosses one threshold ((b) and (c) in figure 7) or both thresholds ((d) in figure 7) they will rate down for imprecision.

Rating down once or twice for imprecision

As the CI gets wider, Core GRADE users will become progressively more uncertain about whether the truth is consistent with an important or unimportant effect, or whether it reflects a non-null effect. To reflect the degree of uncertainty influenced by imprecision of evidence, Core GRADE users can consider rating down one or two levels for imprecision.

A role for plain language statements in making decisions

Stating results in plain language that both clinicians and patients will easily understand is important in making Core GRADE optimally useful for clinical practice. GRADE has therefore provided guidance in making such statements (table 1).[13] We will return to these statements in the sixth paper in this series, in which we discuss Core GRADE summary of findings tables; we introduce them here because they can help decide on rating down once or twice for imprecision.[8]

Rating down once or twice for imprecision: general principles

When deciding whether to rate down twice, two things are worth considering. The first is whether the CI crosses more than one threshold (eg, includes both important benefit and important harm). The second, considering Core GRADE's plain language, is whether the most appropriate message that a particular effect likely exists or that it may exist.

Consider a systematic review comparing reduced versus standard dose corticosteroids for patients with vasculitis. For the outcome of mortality, the authors report a reduction in deaths of 21 per 1000 and a 95% CI that includes a 60 per 1000 reduction but also a 36 per 1000 increase (fig 8). If the authors used an MID of 1% they would rate their certainty in an important effect. Given that the CI crosses the MID threshold they would rate down for imprecision.

Moreover, the width of this CI would prompt the review team to consider rating down twice for imprecision. Indeed, because the CI not only crosses the MID for benefit but also includes an important harm, they would rate down twice for imprecision. Thus, even before considering any other reason for rating down, the authors have only low certainty evidence that the lower dose regimen results in an

BMJ: first published as 10.1136/bmj-2024-081904 on 29 April 2025. Downloaded from https://www.bmj.com/ on 16 August 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Table 1 | GRADE plain language statements when using the null effect or MID thresholds

	Plain language summary	
Certainty	Null effect as threshold	MID as threshold
High	Treatment has a benefit, or Treatment improves outcome X	Treatment has an important benefit, or Treatment has little to no benefit
Moderate	Treatment likely has a benefit, or Treatment likely improves outcome X	Treatment likely has an important benefit, or Treatment likely has little to no benefit
Low	Treatment may have a benefit, or Treatment may improve outcome X	Treatment may have an important benefit, or Treatment may have little to no benefit
Very low	We are very uncertain about whether treatment has a benefit	

The plain language summary pertains to both beneficial and harmful outcomes. Benefit was chosen here for illustration.
GRADE=Grading of Recommendations Assessment, Development and Evaluation; MID=minimal important difference.

important reduction in mortality (the target of their certainty rating).

The second consideration that bears on the decision about rating down once or twice has to do with the most suitable plain language statement that accompanies the certainty of evidence. Consider the example of corticosteroids for patients with vasculitis that includes a CI ranging from a 6% reduction to a 3.6% increase. Assuming systematic review authors do not have concerns on the other Core GRADE domains, would it convey the optimal message about certainty stating that the lower dose regimen likely results in an important reduction in mortality (the statement that would accompany rating down once) or that it may result in an important reduction in mortality (the statement that would accompany rating down twice). If the review authors considered the latter statement more appropriate (as in our view they should) they would rate down twice for imprecision. This highlights that it can be useful for Core GRADE users to consider what would be the most appropriate statement to communicate to their target audience.

The two considerations also apply to imprecision judgments when Core GRADE users choose the null as the threshold of interest. For example, consider a situation in which users rate their certainty in a benefit (threshold the null) but the CI also includes clearly important harm. The finding that the CI is consistent with both benefit and important harm motivates a plain language summary stating that the intervention "may" result in a benefit, and rating down two levels for imprecision.

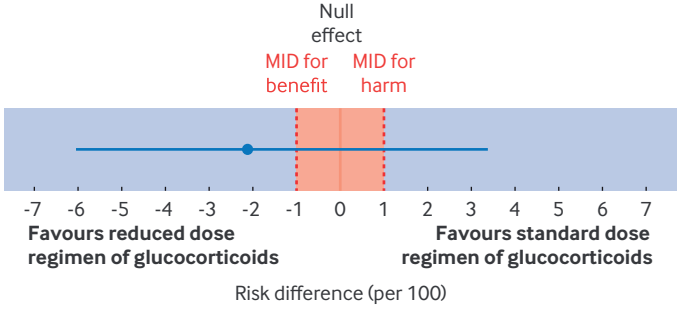

Fig 8 | An example of using Core GRADE to rate down two levels for imprecision in a systematic review of different doses of corticosteroids on mortality in patients with vasculitis. Since the confidence interval includes both important benefit and important harm, the review authors should consider rating down two levels for imprecision. GRADE=Grading of Recommendations Assessment, Development and Evaluation

Rating down for imprecision when effects are large and sample size limited

When the CI crosses the threshold or thresholds of interest, Core GRADE users will rate down for imprecision and do not need to consider sample size. If the CI does not cross the threshold, however, and the effect is large, they must be aware that large effects are unusual in interventions tested in randomised controlled trials. Attempts to replicate results of early studies suggesting such effects often fail. Thus, we suggest that when the CI does not cross the threshold or thresholds of interest and effects on binary outcomes are implausibly large (certainly relative risk reduction >40%, possibly >30%), Core GRADE users should consider rating down for imprecision if the sample size and number of events across all contributing studies are limited.

Our criteria for "limited" rely on routine sample size calculations that would be undertaken when planning a single randomised controlled trial (see supplementary appendix 5). For binary outcomes, these involve specifying the acceptable error rates: α (typically 0.05) and β (typically 0.20), the control group event rate (chosen from the context), and a modest relative risk reduction, typically 20% or 25%. We call the sample size that emerges from the calculation the optimal information size (OIS). If the total sample size of all the studies included in a meta-analysis exceeds the OIS, one does not rate down; if the total sample size proves less that the OIS, one rates down for imprecision. Core GRADE users can consult one of many online calculators to calculate a particular OIS (eg, https://www.openepi.com/SampleSize/SSCohort.htm).

Core GRADE users can make the same calculation for continuous variables by specifying the smallest difference between intervention and control that one would want to avoid missing (ie, the MID), and using the standard deviation from one of the existing studies. An alternative, a rule of thumb, would suggest that to not have concerns about imprecision (ie, to not rate down) would require 400 patients per group (total sample size 800). A previous GRADE article[14] and supplementary appendix 5 provide further details and examples of OIS exploration for both binary and continuous variables.

BMJ: first published as 10.1136/bmj-2024-081904 on 29 April 2025. Downloaded from https://www.bmj.com/ on 16 August 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

Conclusion

The process of assessing the certainty of evidence requires choosing a threshold, either the null or the MID, and then choosing the target of certainty by noting the location of the point estimate in relation to the threshold. When the initial choice of threshold is the null, if the point estimate is close to this threshold, Core GRADE users rate certainty in little to no effect. For judging imprecision, if the CI does not cross the threshold, Core GRADE users typically do not rate down for imprecision; if it crosses the threshold, they do. Core GRADE users may rate down twice when the CI crosses more than one threshold, in particular when it crosses thresholds of important benefit and important harm. Finally, when the CI does not cross the threshold but the effect is large, Core Grade users invoke the OIS and rate down for imprecision if the total sample size fails to meet the OIS criterion.

AUTHOR AFFILIATIONS

[1]Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada

[2]Department of Medicine, McMaster University, Hamilton, ON, L8S 4L8, Canada

[3]MAGIC Evidence Ecosystem Foundation, Oslo, Norway

[4]Pharmacy Department/Evidence-based Pharmacy Centre/Children's Medicine Key Laboratory of Sichuan Province, West China Second University Hospital, Sichuan University, Chengdu, China

[5]Sichuan University and Key Laboratory of Birth Defects and Related Disease of Women and Children, Ministry of Education, Chengdu, China

[6]West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, China

[7]Clinical Research and Epidemiology, Institute of Liver and Biliary Sciences, New Delhi, India

[8]Guide2Guidance, Lemelerberg 7, Utrecht, Netherlands

[9]Evidence-based Practice Center, Mayo Clinic, Rochester, MN, USA

[10]Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore

[11]Division General Internal Medicine, University Hospitals of Geneva, University of Geneva, Geneva, Switzerland

[12]Liverpool School of Tropical Medicine, Liverpool, UK

[13]Department of Medicine, University of Kansas Medical Center, Kansas City, MO, USA

[14]Institute of Health and Society, University of Oslo Faculty of Medicine, Oslo, Norway

[15]Peninsula Dental School, University of Plymouth, Plymouth, UK

[16]Faculty of Dentistry, Manipal University College Malaysia, Malaysia

[17]Department of Cardiology, Shandong Provincial Hospital Affiliated to Shandong First Medical University, Jinan, Shandong, China

[18]Division of Endocrinology, Department of Medicine, Mayo Clinic, Rochester, MN, USA

[19]Knowledge and Evaluation Research Unit, Mayo Clinic, Rochester, MN, USA

[20]HTA Region Stockholm, Centre for Health Economics, Informatics and Health Care Research (CHIS), Stockholm Health Care Services, Stockholm, Sweden

[21]Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden

*Both authors contributed equally (joint first authors)

Contributors: GG and LZeng contributed equally and are joint first authors. GG, VMM, TA, MH, and AI conceived and designed the Core GRADE series. GG, LZeng, MH, RB-P, MP, and HDB drafted this article. All authors critically revised the article across several iterations for important intellectual content and gave final approval for the article. GG and LZeng are the guarantors. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: No external funding.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Provenance and peer review: Not commissioned; externally peer reviewed.

1 Gyuatt G, Agoritsas T, Brignardello-Petersen R, et al. Core GRADE 1: overview of the Core GRADE approach. *BMJ* 2025;389:e081903.
2 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. doi:10.1016/j.jclinepi.2017.05.006
3 Zeng L, Brignardello-Petersen R, Hultcrantz M, et al. GRADE guidelines 32: GRADE offers guidance on choosing targets of GRADE certainty of evidence ratings. *J Clin Epidemiol* 2021;137:163-75. doi:10.1016/j.jclinepi.2021.03.026
4 Zeng L, Brignardello-Petersen R, Hultcrantz M, et al. GRADE Guidance 34: update on rating imprecision using a minimally contextualized approach. *J Clin Epidemiol* 2022;150:216-24. doi:10.1016/j.jclinepi.2022.07.014
5 Jaeschke R, Singer J, Guyatt GH. Measurement of health status. Ascertaining the minimal clinically important difference. *Control Clin Trials* 1989;10:407-15. doi:10.1016/0197-2456(89)90005-6
6 Schünemann HJ, Neumann I, Hultcrantz M, et al, GRADE Working Group. GRADE guidance 35: update on rating imprecision for assessing contextualized certainty of evidence and making decisions. *J Clin Epidemiol* 2022;150:225-42. doi:10.1016/j.jclinepi.2022.07.015
7 Guyatt G, Yao L, Murad MH, et al. Core GRADE 6: presenting the evidence in summary of findings tables. *BMJ* 2025;389:e083866
8 Carrasco-Labra A, Devji T, Qasim A, et al. Minimal important difference estimates for patient-reported outcomes: A systematic survey. *J Clin Epidemiol* 2021;133:61-71. doi:10.1016/j.jclinepi.2020.11.024
9 Zeng L, Yao L, Wang Y, et al. Presentation approaches for enhancing interpretability of patient-reported outcomes in meta-analyses: a systematic survey of Cochrane reviews. *J Clin Epidemiol* 2023;158:119-26. doi:10.1016/j.jclinepi.2023.03.027
10 Zeng L, Li SA, Yang M, et al. Qualitative study of guideline panelists: innovative surveys provided valuable insights regarding patient values and preferences. *J Clin Epidemiol* 2023;161:173-80. doi:10.1016/j.jclinepi.2023.07.014
11 Zeng L, Hultcrantz M, Tovey D, et al. Rating certainty when the target threshold is the null and the point estimate is close to the null. *BMJ Evid Based Med* 2025:bmjebm-2024-113077. doi:10.1136/bmjebm-2024-113077doi:10.1136/bmjebm-2024-113077
12 Stern A, Skalsky K, Avni T, Carrara E, Leibovici L, Paul M. Corticosteroids for pneumonia. *Cochrane Database Syst Rev* 2017;12:CD007720. doi:10.1002/14651858.CD007720.pub3
13 Santesso N, Glenton C, Dahm P, et al, GRADE Working Group. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. *J Clin Epidemiol* 2020;119:126-35. doi:10.1016/j.jclinepi.2019.10.014
14 Guyatt GH, Oxman AD, Kunz R, et al. GRADE guidelines 6. Rating the quality of evidence--imprecision. *J Clin Epidemiol* 2011;64:1283-93. doi:10.1016/j.jclinepi.2011.01.012

Supplementary information: Appendix 1
Supplementary information: Appendix 2
Supplementary information: Appendix 3
Supplementary information: Appendix 4
Supplementary information: Appendix 5

BMJ: first published as 10.1136/bmj-2024-081904 on 29 April 2025. Downloaded from https://www.bmj.com/ on 16 August 2025 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.