



Core GRADE 6: presenting the evidence in summary of findings tables

Gordon Guyatt,^{1,2,3} Liang Yao,⁴ M Hassan Murad,⁵ Monica Hultcrantz,^{6,7} Thomas Agoritsas,^{1,3,8} Hans De Beer,⁹ Stefan Schandelmaier,^{10,11,12} Alfonso Iorio,^{1,2} Linan Zeng,^{13,14,15} Manya Prasad,¹⁶ Per Olav Vandvik,^{3,17} Reem A Mustafa,^{1,18} Arnav Agarwal,^{1,2,3} Tahira Devji,² Ivan D Florez,^{19,20,21} Benjamin Djulbegovic,²² Derek K Chu,^{1,2} Bram Rochwerf,^{1,2} Victor M Montori,^{23,24} Romina Brignardello-Petersen¹

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:e083866
<http://dx.doi.org/10.1136/bmj-2024-083866>

Accepted: 21 March 2025

This sixth article in a seven part series presents the Core GRADE (Grading of Recommendations Assessment, Development and Evaluation) approach to summary of findings tables. These tables provide essential information about the effects of interventions on patient important outcomes, including relative and absolute effects, certainty of evidence, and a plain language summary. For binary outcomes calculating absolute effects requires applying relative risk estimates to baseline risks from studies representative of the target population. For groups of patients with very different baseline risks, summary of findings tables include separate rows with different estimates of absolute effects. For continuous outcomes, challenges arise when individual studies use different instruments to

measure patient reported outcomes. Facilitating interpretation then requires providing details about units of measurement and minimally important differences.

This is the sixth paper in a series describing Core GRADE (Grading of Recommendations Assessment, Development and Evaluation), the essentials of applying GRADE to the conduct and evaluation of systematic reviews, clinical practice guidelines, and health technology assessments. The first article in the series presented an overview of the Core GRADE process for assessing the certainty of evidence and rating the strength of recommendations for clinical questions concerning interventions.¹ The subsequent four articles described how to choose the target of certainty and the domains for rating down and rating up the certainty of the evidence.²⁻⁵ In this article, we describe how Core GRADE users can, using summary of findings tables, provide estimates of effects in relative, absolute, and continuous outcomes and then summarise the results using structured evidence summaries.

Summary of findings tables represent a key GRADE innovation providing an optimal structure to present benefits and harms that clinicians, patients, and decision makers require to guide their choices. Summary of findings tables systematically present numerical results, including the relative and absolute effects, that show the impact of alternative interventions on prioritised patient important outcomes, and the associated certainty of evidence. In this paper, we first describe the components of summary of findings tables and then present considerations for calculating relative and absolute effects for binary outcomes and issues related to presentation of continuous outcomes. We then address problems related to choice of outcomes, what to do when one is unable to pool data through meta-analysis, questions informed by more than one source of evidence, and use of online software for the production and dissemination of summary of findings tables.

The information in this article will enable Core GRADE users to understand the key components in summary of finding tables, including relative and absolute effects, the certainty of evidence, plain language summaries, and reasons for rating certainty

SUMMARY POINTS

Summary of findings tables present the key results of a systematic review in a transparent, structured tabular format, including benefits and harms for outcomes that are important to patients

The tables present both relative and absolute effects, the certainty of evidence for each outcome, and the reasons for rating up or down

When subgroups show credible effect modification, summary of findings tables may include separate rows for different subgroups

When presenting results of binary outcomes, absolute effects are typically derived by applying relative effect estimates to baseline risks from studies representative of the target population

When studies use different instruments to measure continuous outcomes, Core GRADE (Grading of Recommendations Assessment, Development and Evaluation) users can transform data for presentation to the units of a single preferred instrument

Summary of findings tables should include plain language summaries, making it easier for patients, clinicians, and other interest groups to interpret the evidence

Table 1 | Summary of findings table comparing effects of intensive versus less intensive antileukaemic treatment in older adults with acute myeloid leukaemia. Adapted from Sekeres et al⁶

Outcome and follow-up	No of participants (No of studies and type)	Relative effect (95% CI)	Absolute effects (95% CI)		Difference	Certainty of evidence (quality of evidence)	Plain language summary
			Less intensive antileukaemic treatment	Intensive antileukaemic treatment			
Mortality Risk of death Follow-up: longest, range 7.7-60 months	1821 (11 non-randomised studies)	Hazard ratio 0.78 (0.69 to 0.89)	One year risk of dying 578 per 1000*	490 per 1000	88 fewer per 1000 (129 fewer to 42 fewer)	Low† Due to non-randomised studies	Intensive antileukaemic treatment may decrease mortality
Mortality Proportion of people who died Follow-up: one year	87 (one randomised controlled trial)	Risk ratio 0.83 (0.61 to 1.13)	558 per 1000	463 per 1000	95 fewer per 1000 (218 fewer to 73 more)	Low Due to very serious imprecision†	Intensive antileukaemic treatment may decrease mortality
Complete remission duration in months (or time to relapse) Follow-up: longest, range 13.3-60 months	593 (four non-randomised studies)	Four studies reported this outcome. three of the four reported a shorter remission with intensive treatment. The difference in duration ranged from 3.1 months shorter to 0.03 months longer when patients received more intensive versus less intensive treatment				Very low Due to serious imprecision†§	We are very uncertain about the effect of intensive antileukaemic treatment on complete remission
Serious adverse events Proportion of people who had treatment emergent adverse events Follow-up: longest, median 5 years	190 (one non-randomised study)	Risk ratio 1.34 (1.03 to 1.75)	463 per 1000	621 per 1000	157 more per 1000 (14 more to 347 more)	Very low Due to serious risk of bias and imprecision†¶***	We are very uncertain about the effect of intensive antileukaemic treatment on serious adverse events
Duration of hospital stay Follow-up: longest, range 20-60 months	370 (three non-randomised studies)	—	24.36 days	37.32 days	12.96 more days (16.23 fewer to 42.15 more)	Very low Due to very serious inconsistency and very serious imprecision† †† ‡‡	We are very uncertain about the effect of intensive antileukaemic treatment on duration of hospital stay
Quality of life impairment—not reported	—	—	—	—	—	—	—

CI=confidence interval; MID=minimal important difference.

*Event rate from one year mortality was used for less intensive treatment (from non-randomised studies).

†Body of evidence is from non-randomised studies—assessment of certainty of evidence started at low.

‡Using an MID of 50 per 1000, the 95% CI suggests the possibility of an important benefit and an important harm, thus rating down two levels for imprecision.

§Meta-analysis was not possible, but it is likely that the pooled estimate would have crossed the null.

¶The study was at serious risk of bias due to confounding and at moderate risk of bias due to deviation from intended interventions, thus rating down two levels for risk of bias.

***Using an MID of 10%, the 95% CI suggests the possibility of trivial and important harm.

††Not all CIs of the studies overlap, and the statistical heterogeneity is high ($I^2=99\%$).

‡‡Using an MID of five days, the 95% CI suggests important benefit and important harm.

of evidence up or rating down; calculate absolute effects for binary outcomes by applying relative risk estimates to baseline risk; choose the most appropriate presentation format for pooled effect measures when individual studies use different instruments to measure the same continuous outcome; and understand considerations in using software for creating summary of findings tables.

Components of summary of findings tables

For each outcome of interest, summary of findings tables summarise the evidence addressing the effects of interventions versus comparators and the certainty of that evidence, as well as reasons for rating down or up (table 1). Each row in the table focuses on a single outcome presented in the first column. Subsequent columns present the number of participants, number and type of studies, relative and absolute effects presented as risk differences, certainty of evidence, and a plain language summary describing the effect. Although the format for presenting summary of findings tables may differ, the key information they present should not. Supplementary appendix 1 shows an alternative format we refer to as an evidence profile.

Assessing the certainty of evidence involves choices and judgment. Those making the judgments must communicate their rationale in succinct, explicit, and transparent footnotes (ie, explanations) with the summary of findings table. This is one of the strengths of Core GRADE: although two Core GRADE users may arrive at different judgments when looking at the same body of evidence, providing sufficient information for readers to understand their rationale will clarify the reasons for disagreement. For instance, they may have chosen different thresholds for the minimal important difference (MID), the smallest magnitude of effect that patients consider important. Understanding these reasons may allow, in different contexts, adaptation of existing summary of findings tables to different choices or judgments. Explanations should be concise, informative, relevant, easy to understand, and accurate.⁷ Supplementary appendix 2 provides more details and examples of how to write optimal explanations.

A previous GRADE paper provided guidance for wording plain language summaries communicating the effects while conveying the assigned level of certainty of the evidence, thus facilitating understanding of evidence summaries.⁸ Box 1 summarises this guidance

Box 1: Writing standardised GRADE plain language summaries in summary of findings tables

Standardised plain language summaries should convey, for each outcome, information about the certainty of the evidence and the effect of the intervention. The following qualifiers then inform the direction of the effect:

- High certainty: reduces, increases, or has little to no effect
- Moderate certainty: probably (likely) reduces, increases, or has little to no effect
- Low certainty: may (possibly) reduce, increase, or have little to no effect
- Very low certainty: the evidence is very uncertain; or the effect is very uncertain

When focusing on the target of certainty in relation to the null, plain language summaries should communicate that there is a benefit or harm, which is to be understood as a non-null effect. Examples:

- High certainty: knee arthroscopy increases function
- Moderate certainty: knee arthroscopy probably (likely) increases function
- Low certainty: knee arthroscopy may (possibly) increase function
- Very low certainty: the effect of knee arthroscopy on function is very uncertain

When focusing on the minimal important difference, plain language summaries should communicate that there is an important benefit or harm, or alternatively that there is little to no important effect. Examples:

- High certainty of an important effect: knee arthroscopy results in an important increase in function
- Moderate certainty of little to no effect: knee arthroscopy probably has little to no important effect on function
- Low certainty of an important effect: knee arthroscopy may (possibly) result in an important increase in function
- Very low certainty: the effect of knee arthroscopy on function is very uncertain

GRADE=Grading of Recommendations Assessment, Development and Evaluation.

as well as additional guidance related to the null and MID thresholds that are the focus of Core GRADE.

Presenting dichotomous outcomes**Presenting relative measures of effect**

We have emphasised in both the first and the second article in this series dealing with establishing the target of certainty rating and rating down for imprecision¹² that relative risks are usually similar across different baseline risks.⁹⁻¹³ Thus, systematic review authors will usually conduct meta-analyses of relative effects, using either relative risks, odds ratios, or hazard ratios. Relative risks are easier to interpret than odds ratios and facilitate estimation of absolute effects, making it the preferred measure from a user's perspective.¹⁴ The odds ratio, however, can address problems that occur with relative risks when baseline risks are high (>50%).¹⁵ Such situations are unusual and relative risks usually serve well.¹⁶

One can interpret hazard ratios, which consider not only if an event occurs but when it occurs, similarly to relative risks.¹⁷ Investigators often use hazard ratios when mortality is the outcome of interest and death rates are high, as in the specialty of oncology (table 1, first and second row). Under these circumstances, duration of survival becomes the most important factor for patients.¹⁶

Because patients and key stakeholders are primarily concerned about absolute rather than relative effects, summary of findings tables include absolute

measures—risk differences—for key binary outcomes. To generate risk differences, they apply relative estimates that come from the meta-analyses to baseline risks. In making their decisions, they consider the magnitude and associated certainty of those absolute effects. The next section presents the method of calculating and how to present absolute effects.

Calculating and presenting absolute measures of effect*Applying relative effects to baseline risks*

Because they depend on baseline risks, absolute effects vary from one clinical scenario to another. Therefore, the first step for Core GRADE users when calculating absolute effects is to select a specific baseline risk for the patient group under consideration. In doing so they should decide on a time frame for measuring outcomes and in general use that same time frame for all outcomes.

Use of the same time frame for all outcomes is not an absolute rule. For example, it is not unusual for benefits to occur over a short time frame but for rare serious events to occur only over the long term. A summary of findings table needs to report both outcomes using the appropriate time frame.

Because of the often selective sampling process in many randomised trials, the baseline risk will ideally come from a rigorous, large observational study that includes a more generalisable population, a systematic review of such studies, or a large pragmatic trial with broad eligibility criteria. Such studies are often unavailable and as a result systematic review authors will often use the median event rate in the comparator arms across all randomised trials included in the review.^{16 18}

To obtain risk differences, authors apply pooled relative effects to chosen baseline risks. To illustrate with an example, consider the outcome mortality as calculated from randomised controlled trials—row 2 in table 1. The body of evidence (in this case from a single randomised trial) suggests that, when comparing intensive versus less intensive antileukaemic treatment, the relative risk is 0.83 (a 17% relative risk reduction). Applying this relative risk reduction to the baseline risk of death among older adults who receive less intensive antileukaemic treatment (56%), the absolute risk reduction with intensive antileukaemic treatment is calculated as 17% multiplied by 56%, equaling a 9.5% absolute risk reduction. Supplementary appendix 3 provides more details on how to make this calculation.

Different risk groups

Clinicians can sometimes, considering prognostic factors for outcomes of interest, identify patients with sufficiently different baseline risks to warrant different management strategies. When this is the case, Core GRADE users may present separate risk differences for patient groups at varying risk of events, which can lead to different decisions.

Table 2 presents an example from a systematic review supporting the development of recommendations

Table 2 | Summary of findings table comparing effects of SGLT-2 inhibitors versus standard care in people with type 2 diabetes. Adapted from Li et al¹⁹ and Palmer et al²⁰ to show only outcomes of interest

Outcome Time frame five years	Study results and measurements	Absolute effect estimates		Certainty of evidence (quality of evidence)	Plain language summary
		Standard care	SGLT-2 inhibitors		
All cause mortality (very low risk: adults with ≤ 3 risk factors and no underlying cardiovascular disease or chronic kidney disease)	Odds ratio 0.77 (95% CI 0.71 to 0.83) Based on data from 282 704 participants in 225 studies	20 per 1000 Difference 5 fewer per 1000 (95% CI 6 fewer to 3 fewer)	15 per 1000	High	There is no important difference between SGLT-2 inhibitors and standard care
All cause mortality (very high risk: patients with cardiovascular disease and chronic kidney disease)	Odds ratio 0.77 (95% CI 0.71 to 0.83) Based on data from 282 704 participants in 225 studies	265 per 1000 Difference 48 fewer per 1000 (95% CI 61 fewer to 35 fewer)	217 per 1000	High	SGLT-2 inhibitors reduce the risk of death compared with standard care

CI=confidence interval; SGLT-2=sodium-glucose cotransporter-2.

regarding the use of sodium-glucose cotransporter-2 (SGLT-2) inhibitors versus standard care for people with type 2 diabetes.^{19 20} The guideline authors, aware of the large gradients in risk of death in people with type 2 diabetes depending on their risk factors and cardiovascular and renal morbidities, made recommendations taking into account these baseline risks. The guideline panel identified five risk strata and ultimately made recommendations specific to each stratum. Here, we highlight results for the most important outcome, all cause mortality, from the two extremes: people at lowest risk and those at highest risk.

The authors began by conducting a meta-analysis to calculate the relative effects of SGLT-2 inhibitors on mortality across the entire population and reported an odds ratio of 0.77 with a 95% confidence interval (CI) from 0.71 to 0.83 (table 2). The authors explored possible subgroup differences in relative effects across risk groups. They found no evidence of effect modification, including at the extremes of baseline risk. They therefore applied the relative risk to the baseline risks in the two risk groups over a period of five years (20 per 1000 deaths in people at very low risk; 265 per 1000 in the highest risk group). The result was a risk difference of 5 fewer per 1000 with SGLT-2 inhibitors in the lowest risk group and 48 fewer per 1000 in the highest risk group. The synthesis team presented the evidence for all five subgroups in the summary of findings in adjacent rows—table 2 does so for the highest and lowest risk groups.

Considering not only mortality but other relevant outcomes, the panel made a weak recommendation against treatment in the very low risk group and a strong recommendation in favour in the very high risk group. In other situations, particularly when displaying multiple important outcomes, it may be appropriate to develop separate summary of findings tables for each risk group.

Using different risk groups may lead to different targets of certainty rating, as it did here (target little or no effect in the low risk group and an important effect in the high risk group). This may then result in different decisions about precision and inconsistency. Here, using an MID in mortality of 10 in 1000, the entire CI for the low risk group fell in the range of no important

effect and in the high risk group fell entirely in the range of an important effect, in both cases indicating no serious imprecision.

Directly calculating risk differences

Calculating risk differences by applying relative effects to baseline risks may, in some scenarios, result in misleading point estimates of effect and even more misleading and asymmetrical CIs.²¹ Such results often occur when the outcome is rare (event rates <2% and most problematic <1%). Faced with this problem, rather than conducting meta-analysis of relative effects, review authors should generally conduct meta-analyses of risk differences. Supplementary appendix 4 provides examples.

Presenting continuous outcomes

When studies use the same measure

In many cases, studies reporting an outcome measured as a continuous variable use the same instrument or scale across studies. Take, for instance, length of hospital stay (measured in days) in table 1 or pain (often measured using a 10 cm visual analogue scale) in table 3.

When studies use the same measure, an intuitive measure of effect is the mean difference. Because natural units (that is, the units of the outcome measure, for instance hospital days, or, in table 3, option 1, the units of the KOOS (Knee injury and Osteoarthritis Outcome Score) pain scale rather than statistical units such as the standardised mean difference) are used, Core GRADE users present the mean difference in the absolute effects column. To facilitate interpretation, the preferred approach is to provide information about the outcome in the comparison group (for example, in table 1 the median duration of hospital stay in the control arm was 24 days), the intervention group (37 days in table 1), and the difference between the two (13 days).

In addition to describing the time point of interest and method of measurement of the outcomes (eg, visual analogue scale, KOOS pain scale, days in hospital), the summary of findings table should provide, when it is not obvious, the range and direction of the scale or the units in which the results are presented. For example, blood loss may be presented in millilitres or ounces,

Table 3 | Summary of findings table comparing effects of knee arthroscopy versus conservative management in people with degenerative knee disease. Adapted from Brignardello-Petersen et al²² and Siemieniuk et al²³ to show only pain outcomes measured by different instruments

Outcome Time frame	Studies and measurements	Absolute effect estimates		Certainty of evidence (quality of evidence)	Plain language summary
		Conservative management	Arthroscopy		
Option 1 Pain (difference in change from baseline) Three months	Measured by different instruments converted to scale of index instrument (KOOS pain subscale—MID 12) Scale 0–100, higher scores better Based on data from 1231 participants in 10 studies Follow-up three months	Mean 15.00 points Mean difference 5.38 more (95% CI 1.95 more to 8.81 more)	Mean 20.00 points	Moderate*† Different pain summary measures suggest different effects	Knee arthroscopy probably does not result in an important reduction in pain
Option 2 Pain (difference in proportion of patients who achieve a change greater than the MID) Three months	Based on data from 1102 participants in nine studies Follow-up three months‡	669 per 1000§ Difference 124 more per 1000 (95% CI 44 more to 204 more)	793 per 1000	Low†¶ Serious imprecision and different pain summary measures suggest different effects	Knee arthroscopy possibly increases the number of patients with an important reduction in pain
Option 3 Pain (SMD in change from baseline) Three months	Measured by different instruments with different scales Scale: Higher scores better (either in original instruments or converted) Based on data from 1231 participants in 10 studies Follow-up three months	Mean 15.00 SMD 0.16 higher (95% CI 0.03 higher to 0.28 higher)	Mean 17.04	Low†** Serious imprecision and different pain summary measures show different effects	Knee arthroscopy possibly does not result in an important reduction in pain

CI=confidence interval; KOOS=Knee injury and Osteoarthritis Outcome Score; MID=minimal important difference; SMD=standardised mean difference.
 *Although the I² statistic was high, all point estimates suggested little to no effect and therefore no rating down for inconsistency.
 †Some pain summaries suggest small but important differences in pain, whereas others suggest only unimportant differences, warranting rating down for inconsistency.
 ‡Because one study provided data allowing calculation of the proportion of patients who achieved a change greater than the MID, 10 studies were relevant for options 1 and 3 and nine for option 2.
 §Baseline risk of 669 per 1000 based on the median of number of people who achieved an improvement of MID or greater in the conservative management groups.
 ¶Using an MID of 100 per 1000 patients (10%), the 95% CI crosses this threshold, suggesting the possibility of a trivial benefit and thus warranting rating down for inconsistency.
 **Using Cohen's threshold for interpretation of SMD, the 95% CI crosses the MID (small effect threshold) of 0.2, suggesting the possibility of an important benefit and thus warrants rating down for imprecision.

and visual analogue scales for measuring pain are used with ranges of 0–10 cm or 0–100 mL—the interpretation of the same mean difference of 2 points will be very different if a 10 point scale versus a 100 point scale is used. Similarly, some patient reported outcomes might be measured with instruments in which higher scores signify better outcomes, and others in which lower scores signify better outcomes (we present an example in the next section). Core GRADE users should therefore communicate the direction of the instrument's scale.

Finally, perhaps the most useful way of ensuring the interpretability of the outcome is to designate the smallest difference patients perceive as important, the MID. For instance, a footnote in table 1 specifies the MID for reduction in hospital stay of five days and the first row in table 3 specifies the MID for the KOOS pain scale of 12.

When studies use different measures

Researchers sometimes measure the same outcome using different instruments. This most often occurs in health status measures that address constructs such as health related quality of life, function, or severity of symptoms. Multiple instruments are often available, and investigators make different choices for their studies. When this occurs, it presents challenges for systematic review authors.

Table 3 presents an example from a systematic review assessing the effect of knee arthroscopy versus conservative management in people with degenerative knee disease²² developed to support a guideline.²³ Across the included randomised trials, researchers used seven different instruments to measure the

outcome of function related to knee arthritis.^{24–30} When preparing their summary of findings table, Core GRADE users have several options for summary effect measures,^{31–33} three of which are illustrated in table 3. A previous GRADE paper provides additional details of the alternatives,³² and another previous paper details the statistical methods involved.³¹

Option 1: Using the mean difference on the scale of an index instrument

When one of the instruments is more commonly used and understandable than others (the index instrument), Core GRADE users can convert the other scores into the scale of the index instrument. They then conduct the meta-analysis using the mean difference on the index instrument as the measure of effect. In the example, clinical experts suggested KOOS, which uses a scale of 0–100 (with higher scores representing better function), as the index instrument.

When studies reported function using the Western Ontario and McMaster Universities' arthritis index domain (range from 0 to 68, with higher scores representing worse function) transforming to a scale of 0–100 and reversing the direction allowed combining results across studies using the mean difference (table 3, first row). When choosing this option, systematic reviewers are likely to require different conversions for each instrument (for example, scores from the Arthritis Impact Measurement Scale, with a range 0–10, multiplied by 10) or may need no conversion at all if another instrument uses the same range (for example, SF-36 survey scores also have a scale from 0 to 100 in the same direction as KOOS).

BMJ: first published as 10.1136/bmj-2024-083866 on 27 May 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.

Authors can then make the results on the index instrument interpretable by relating it to the MID. In this case, the best estimate of the KOOS MID is 12 units.³⁴ Because the mean difference (5.38) is much less than the MID, one can conclude that any improvement as a result of the intervention is small and likely unimportant.

Option 2: Presenting data as binary outcomes

If systematic reviewers or guideline developers know what the MID is for each of the instruments and assume a normal distribution of results, they can calculate the proportion of people who experience an improvement larger than the MID within each arm, thereby obtaining a risk ratio or risk difference for each of the studies. They can then pool these proportions across studies. Alternatively, if an MID for each instrument is not available, reviewers can convert scores into the scale of the index instrument with a known MID. Based on the results of a systematic review of the literature addressing MIDs across instruments,³⁵ the second row of table 3 shows the results of such an analysis. Previous papers describe the underlying statistical methods.^{16 31}

Option 3: Calculating the standardised mean difference

When included studies assess the same outcome but measure it using a variety of different scales review authors sometimes choose another option, the standardised mean difference (SMD) as a summary statistic.³⁶ By dividing the mean difference between the treatment and control groups by the pooled sample standard deviation in each study at a specific time point, the SMD converts the results of all studies to standard deviation units.³⁷

The SMD is prone to two major problems. Firstly, SMDs express the size of the treatment effect in each study relative to the variability observed in that study. Many may find the resulting treatment effect reported in standard deviation units rather than the original units of measurement difficult to interpret.³⁷⁻³⁹ Secondly, although guidance is available to interpret SMDs (ie, an SMD of 0.2 is the threshold for a small and important effect),⁴⁰ clinicians may be appropriately sceptical of this threshold, which is limited by large variability in the methods investigators use to calculate the SMD.^{39 41} Lastly, the SMD bears a highly variable relation to the actual magnitude of effect: in two studies with exactly the same magnitude of effect, the SMD will be much larger in a study that enrolls a homogeneous population than in one that enrolls a heterogeneous population.⁴¹

All these options have limitations. Although the SMD (option 3) remains the most used summary statistic, for the reasons we have noted, it is often the least satisfactory. Review authors should consider using the SMD only when the outcome is reported using multiple scales and no instrument measuring the construct has a credible MID available.^{42 43} For many clinicians, using natural units in relation to their MIDs (option 1)

remains an unfamiliar approach, and in interpreting results they must avoid the temptation of inferring that any effect smaller than the MID is unimportant. Converting the continuous to a binary outcome (option 2) is appealing in that this is a presentation with which clinicians are familiar. Other options remain available, including the ratio of means, a presentation that some find intuitive and appealing.⁴⁴

We suggest presenting the mean difference and interpreting these differences in relation to the MID (option 1), as well as the binary outcome approach (option 2).¹⁴ If results are concordant between these approaches, review authors may make strong inferences about the apparent magnitude of effect. If not, inferences about magnitude must be weaker. Indeed, in table 3 the methods that present a continuous variable (options 1 and 3) suggest a small and likely unimportant effect, whereas the binary approach (option 2) provides a point estimate of 12.4% more individuals who gain an important improvement, an effect that most patients are likely to consider important.

Core GRADE users will acknowledge the discrepancy between interpretations of the results and rate down certainty of evidence accordingly. They must still, however, take responsibility for a coherent message for clinicians and patients who rely on them for guidance. Given the evidence in table 3, conclusions of little or no effect or a small but important effect are both reasonable. Guideline panels will need to come down on one side or the other and present results in their summary of findings tables that support their inference. The text discussion should, however, present alternative results in explaining the lower certainty of evidence supporting their ultimate inferences.

Additional considerations for summary of findings tables

Choosing which outcomes to present

Previous GRADE guidance suggested limiting the number of outcomes to seven,¹⁶ but this may not adequately serve the needs of the target audiences of a particular systematic review. For instance, target audiences may be interested in the effect of an intervention at different time points; they may wish to see different ways of presenting the outcome, such as in the example in table 3, including both as a dichotomous and as a continuous outcome; they may wish to see evidence for a single outcome from both non-randomised studies and randomised trials such as for the outcome mortality in table 1; if no evidence is available for a critical or important outcome, Core GRADE users may consider including a row that describes this explicitly (eg, outcome quality of life in row 6 of table 1). For all these reasons, decision makers or clinician audiences may require tables that have more than seven rows.

Thus, when creating summary of findings tables, we suggest being parsimonious but flexible and using online supplementary material and digital publication platforms that allow the publication of interactive or

additional summary of findings tables (see section “Software”). What Core GRADE users should not do is have rows for outcomes that overlap (eg, all cause mortality and cardiovascular mortality) because of the risk of double counting, as occurred in a summary of findings table that included both overall major bleeding and gastrointestinal bleeding.⁴⁵

When data cannot be pooled

While one can, for individual studies, always present quantitative data that authors report, in some instances it is not possible to use a meta-analysis to pool the results across studies. For example, authors may not have reported the data in a way that allows review authors to transform the data to accommodate their analysis, or the reported data may be insufficient. Such circumstances require a narrative synthesis.

In these situations, Core GRADE users will follow the same principles in their data synthesis and presentation as they do when they can generate pooled estimates: they will provide a summary at the outcome level and rate the certainty of the evidence. They will present their narrative summary in the same columns as the effects of interventions. In the absence of a pooled estimate, they can still make judgments on the certainty of the evidence rating using the same GRADE certainty domains.⁴⁶

The third row of table 1 provides an example for the outcome of complete remission, reported in the studies as median rather than mean. In this case, clinical experts advised against converting to means and the studies did not report sufficient data to do a meta-analysis using the difference in medians as the measure of effect. In this case, the summary provided the number of studies with positive and negative results and the range of differences between intervention and comparator in duration of illness across studies. The Cochrane Handbook provides detailed explanations and examples of how to summarise these types of data.⁴⁷

Burden of treatment

In addition to displaying benefits and harms in their summary of findings tables, Core GRADE users may choose to narratively summarise available data on the burden of interventions^{48 49}—what has been called the “work” of being a patient and what GRADE guidelines have referred to as practical issues. Supplementary appendix 5 provides more details and examples on presenting practical issues in summary of findings tables. This work includes drug frequency and route, tests, and clinic visits; procedures and devices; coordination of care; recovery and adaptation; managing dependencies; directions for how patients should manage their diet, exercise, and health habits; adoption and routine use of digital self-management and clinical communication tools; impact on work and social life; and any physical or emotional distress that may come with managing all these issues.⁵⁰ Having these data summarised in summary of findings tables can enable their use in shared decision making with patients and in the tools that support this practice.^{51 52}

Direct versus indirect evidence

When direct evidence for an important outcome is limited, indirect evidence may provide the highest certainty available evidence, and thus inform the summary of findings table. In a previous article in this series addressing indirectness we have described how Core GRADE users might best handle indirect evidence in summary of findings tables.⁵

Two common situations often arise. In the first, authors may need to present syntheses of evidence from a different but related population. For instance, evidence regarding harms of an intervention applied to a rare disease may come from studies of the intervention in other more common conditions. The extent of rating down for indirectness would then depend on the likelihood that adverse effects would be similar across conditions.

In another common scenario, Core GRADE users may need to rely on surrogate outcomes to make inferences about a patient important outcome. As described in the indirectness article,⁵ in such instances the summary of findings table presents inferences about the impact of treatment on the patient important outcome while making clear those inferences are based on results from a surrogate, rating down once or twice for indirectness.¹⁶

When data are available from randomised trials and non-randomised studies

Non-randomised studies of interventions (NRSI) can provide relevant information when synthesising evidence addressing the effects of interventions. Randomised trials may not be available for one or more outcomes for many reasons, one being a harm that occurs infrequently.⁵³

The fourth article in this series provides guidance for assessing risk of bias in NRSI as well as circumstances in which one might rate up certainty of evidence.⁴ Guidance regarding other reasons for rating down, including imprecision, inconsistency, and indirectness for randomised trials also applies to NRSI.

When information from both randomised trials and NRSI exists, Core GRADE users should prioritise the source with the highest certainty. As is the case for indirect evidence, when the certainty of the evidence from randomised trials and non-randomised studies is similar, presenting both bodies of evidence in adjacent rows may be desirable (see table 1, mortality).

Software

When creating their structured summaries of evidence, Core GRADE users may benefit from online software that allows for structuring the data and customising the format of the summary of findings tables while ensuring all key components are included. Such software also facilitates the calculation of absolute effects when presenting binary outcomes, and the creation of plain language summaries. Core GRADE users should bear in mind, however, that while available software embeds GRADE guidance, it will not guide users in making judgments and will not flag

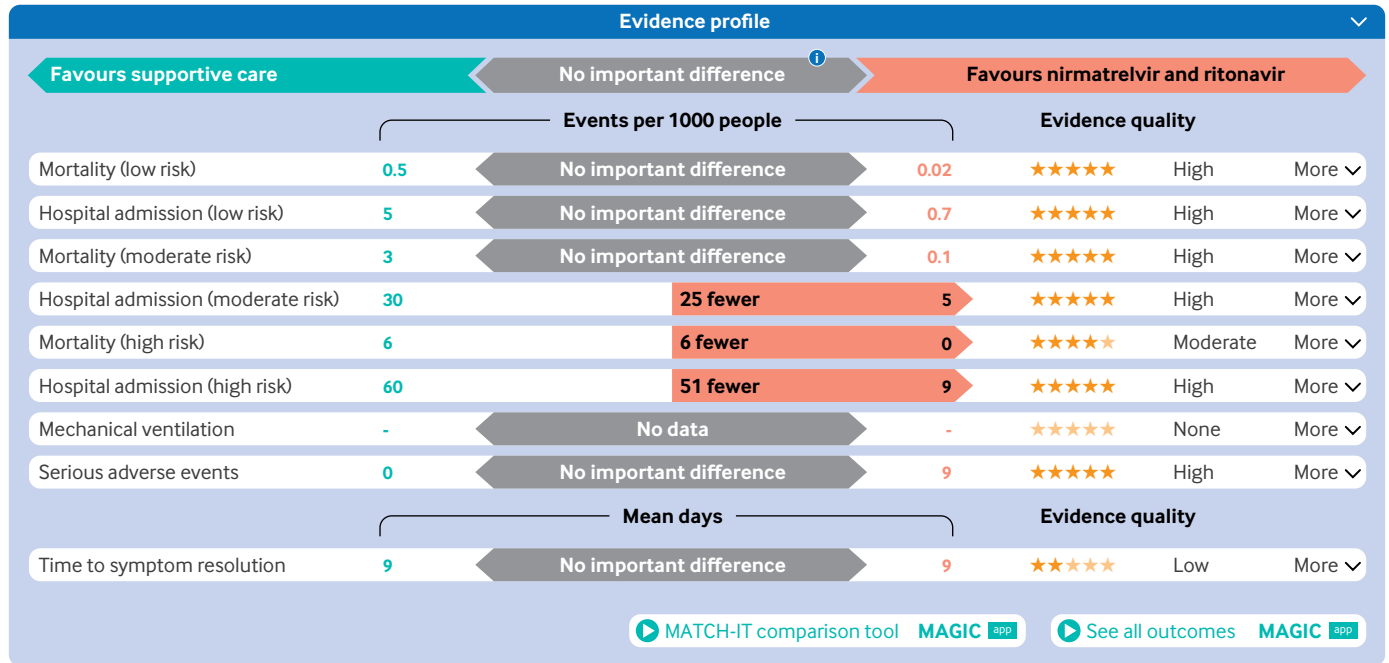


Fig 1 | Example of table presenting the impact of nirmatrelvir and ritonavir on outcomes in patients with non-severe covid-19, with links to tools for shared decision making

potentially challenging situations, It may even suggest judgments that are inaccurate. Examples of potential inaccuracies include automatically rating up for large effect, not flagging situations when it is best to analyse risk differences directly, or when reviewers face close call decisions in more than one domain that warrant a gestalt judgment across the entire body of evidence (a situation described in the first article of this series).¹

MAGICapp (www.magicapp.org) and GRADEpro GDT (www.grade.pro) are two online platforms that enable users to create and export summary of findings tables in their publications, presentations, and teaching materials. Additional paid features allow for digital publication of alternative formats, including interactive formats, multilayered presentations of information, one page presentations, infographics, and mobile apps, as well as dynamic updates of the evidence—for example, within the context of living systematic reviews and guidelines. Figure 1 shows an example of a multilayered table, here presented with an interactive infographic directly linking outcomes to decision aids that support shared decision making (<https://www.bmj.com/content/370/bmj.m3379>).⁵⁴

Conclusions

Optimal summary of findings tables allow users to best understand the results of the synthesis and appraisal of a body of evidence. They are crucial for guideline panels making recommendations and for health technology assessment reports. Using summary of findings tables ensures the explicit and transparent presentation of all relevant information. Summary of findings tables can take various forms, but they

all share the main features and principles that Core GRADE users should follow.

AUTHOR AFFILIATIONS

- ¹Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, ON, Canada
- ²Department of Medicine, McMaster University, Hamilton, ON, L8S 4L8, Canada
- ³MAGIC Evidence Ecosystem Foundation, Oslo, Norway
- ⁴Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore
- ⁵Evidence-based Practice Center, Mayo Clinic, Rochester, MN, USA
- ⁶HTA Region Stockholm, Centre for Health Economics, Informatics and Health Care Research (CHIS), Stockholm Health Care Services, Stockholm, Sweden
- ⁷Department of Learning, Informatics, Management and Ethics, Karolinska Institutet, Stockholm, Sweden
- ⁸Division General Internal Medicine, University Hospitals of Geneva, Geneva, Switzerland
- ⁹Guide2Guidance, Lemelerberg 7, Utrecht, Netherlands
- ¹⁰Division of Clinical Epidemiology, University Hospital and University of Basel, Basel, Switzerland
- ¹¹School of Public Health, University College Cork, Cork, Ireland
- ¹²MTA–PTE Lendület “Momentum” Evidence in Medicine Research Group, Medical School, University of Pécs, Pécs, Hungary
- ¹³Pharmacy Department/Evidence-based Pharmacy Centre/Children’s Medicine Key Laboratory of Sichuan Province, West China Second University Hospital, Sichuan University, Chengdu, China
- ¹⁴Sichuan University and Key Laboratory of Birth Defects and Related Disease of Women and Children, Ministry of Education, Chengdu, China
- ¹⁵West China Biomedical Big Data Center, West China Hospital, Sichuan University, Chengdu, China
- ¹⁶Clinical Research and Epidemiology, Institute of Liver and Biliary Sciences, New Delhi, India
- ¹⁷Institute of Health and Society, University of Oslo Faculty of Medicine, Oslo, Norway
- ¹⁸Department of Medicine, University of Kansas Medical Center, Kansas City, MO, USA

¹⁹Department of Pediatrics, University of Antioquia, Medellín, Colombia

²⁰Pediatric Intensive Care Unit, Clínica Las Américas-AUNA, Medellín, Colombia

²¹School of Rehabilitation Science, McMaster University, Hamilton, ON, Canada

²²Division of Hematology/Oncology, Department of Medicine, Medical University of South Carolina, SC, USA

²³Division of Endocrinology, Department of Medicine, Mayo Clinic, Rochester, MN, USA

²⁴Knowledge and Evaluation Research Unit, Mayo Clinic, Rochester, MN, USA

Contributors: GG, VMM, TA, MH, and AI conceived and designed the Core GRADE series. GG, RB-P, LY, and MHM drafted this article. All authors critically revised the article across several iterations for important intellectual content and gave final approval for the article. GG is the guarantor. 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.

- Guyatt G, Agoritsas T, Brignardello-Petersen R, et al. Core GRADE 1: overview of the Core GRADE approach. *BMJ* 2025;389:e081903.
- Guyatt G, Zeng L, Brignardello-Petersen R, et al. Core GRADE 2: choosing the target of certainty rating and assessing imprecision. *BMJ* 2025;389:e081904.
- Guyatt G, Schandelmaier S, Brignardello-Petersen R, et al. Core GRADE 3: rating certainty of evidence—assessing inconsistency. *BMJ* 2025;389:e081905.
- Guyatt G, Wang Y, Eachempati P, et al. Core GRADE 4: rating certainty of evidence—risk of bias, publication bias, and reasons for rating up certainty. *BMJ* 2025;389:e083864.
- Guyatt G, Iorio A, De Beer H, et al. Core GRADE 5: rating certainty of evidence—assessing indirectness. *BMJ* 2025;389:e083865.
- Sekeres MA, Guyatt G, Abel G, et al. American Society of Hematology 2020 guidelines for treating newly diagnosed acute myeloid leukemia in older adults. *Blood Adv* 2020;4:3528-49. doi:10.1182/bloodadvances.2020001920
- Santesso N, Carrasco-Labra A, Langendam M, et al. Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments. *J Clin Epidemiol* 2016;74:28-39. doi:10.1016/j.jclinepi.2015.12.006
- 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
- Deeks JJ. Issues in the selection of a summary statistic for meta-analysis of clinical trials with binary outcomes. *Stat Med* 2002;21:1575-600. doi:10.1002/sim.1188
- Schmid CH, Lau J, McIntosh MW, Cappelleri JC. An empirical study of the effect of the control rate as a predictor of treatment efficacy in meta-analysis of clinical trials. *Stat Med* 1998;17:1923-42. doi:10.1002/(SICI)1097-0258(19980915)17:17<1923::AID-SIM874>3.0.CO;2-6
- Furukawa TA, Guyatt GH, Griffith LE. Can we individualize the 'number needed to treat'? An empirical study of summary effect measures in meta-analyses. *Int J Epidemiol* 2002;31:72-6. doi:10.1093/ije/31.1.72
- Torres Roldan VD, Ponce OJ, Urtecho M, et al. Understanding treatment-subgroup effect in primary and secondary prevention of cardiovascular disease: An exploration using meta-analyses of individual patient data. *J Clin Epidemiol* 2021;139:160-6. doi:10.1016/j.jclinepi.2021.08.006
- Hanlon P, Butterly EW, Shah AS, et al. Treatment effect modification due to comorbidity: Individual participant data meta-analyses of 120 randomised controlled trials. *PLoS Med* 2023;20:e1004176. doi:10.1371/journal.pmed.1004176
- Johnston BC, Alonso-Coello P, Friedrich JO, et al. Do clinicians understand the size of treatment effects? A randomized survey across 8 countries. *CMAJ* 2016;188:25-32. doi:10.1503/cmaj.150430
- Rochwerf B, Elbarbary M, Jaeschke R, et al. Understanding the Results: More About Odds Ratios. In: Guyatt G, Rennie D, Meade MO, et al, eds. *Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice*. 3rd ed. McGraw-Hill Education, 2015.
- Guyatt GH, Oxman AD, Santesso N, et al. GRADE guidelines: 12. Preparing summary of findings tables-binary outcomes. *J Clin Epidemiol* 2013;66:158-72. doi:10.1016/j.jclinepi.2012.01.012
- Sashegyi A, Ferry D. On the Interpretation of the Hazard Ratio and Communication of Survival Benefit. *Oncologist* 2017;22:484-6. doi:10.1634/theoncologist.2016-0198
- Hemkens LG, Contopoulos-Ioannidis DG, Ioannidis JPA. Routinely collected data and comparative effectiveness evidence: promises and limitations. *CMAJ* 2016;188:E158-64. doi:10.1503/cmaj.150653
- Li S, Vandvik PO, Lytvyn L, et al. SGLT-2 inhibitors or GLP-1 receptor agonists for adults with type 2 diabetes: a clinical practice guideline. *BMJ* 2021;373:n1091. doi:10.1136/bmj.n1091
- Palmer SC, Tendal B, Mustafa RA, et al. Sodium-glucose cotransporter protein-2 (SGLT-2) inhibitors and glucagon-like peptide-1 (GLP-1) receptor agonists for type 2 diabetes: systematic review and network meta-analysis of randomised controlled trials. *BMJ* 2021;372:m4573. doi:10.1136/bmj.m4573
- Murad MH, Montori VM, Walter SD, Guyatt GH. Estimating risk difference from relative association measures in meta-analysis can infrequently pose interpretational challenges. *J Clin Epidemiol* 2009;62:865-7. doi:10.1016/j.jclinepi.2008.11.005
- Brignardello-Petersen R, Guyatt GH, Buchbinder R, et al. Knee arthroscopy versus conservative management in patients with degenerative knee disease: a systematic review. *BMJ Open* 2017;7:e016114. doi:10.1136/bmjopen-2017-016114
- Siemieniuk RAC, Harris IA, Agoritsas T, et al. Arthroscopic surgery for degenerative knee arthritis and meniscal tears: a clinical practice guideline. *BMJ* 2017;357:j1982. doi:10.1136/bmj.j1982
- Meenan RF, Gertman PM, Mason JH. Measuring health status in arthritis. The arthritis impact measurement scales. *Arthritis Rheum* 1980;23:146-52. doi:10.1002/art.1780230203
- Roos EM, Toksvig-Larsen S. Knee injury and Osteoarthritis Outcome Score (KOOS) - validation and comparison to the WOMAC in total knee replacement. *Health Qual Life Outcomes* 2003;1:17. doi:10.1186/1477-7525-1-17
- Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. *J Rheumatol* 1988;15:1833-40.
- Scuderi GR, Bourne RB, Noble PC, Benjamin JB, Lonner JH, Scott WN. The new Knee Society Knee Scoring System. *Clin Orthop Relat Res* 2012;470:3-19. doi:10.1007/s11999-011-2135-0
- Ware JE Jr, Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care* 1992;30:473-83. doi:10.1097/00005650-199206000-00002
- Briggs KK, Lysholm J, Tegner Y, Rodkey WG, Kocher MS, Steadman JR. The reliability, validity, and responsiveness of the Lysholm score and Tegner activity scale for anterior cruciate ligament injuries of the knee: 25 years later. *Am J Sports Med* 2009;37:890-7. doi:10.1177/0363546508330143
- Garratt AM, Brealey S, Gillespie WJ, DAMASK Trial Team. Patient-assessed health instruments for the knee: a structured review. *Rheumatology (Oxford)* 2004;43:1414-23. doi:10.1093/rheumatology/keh362
- Thorlund K, Walter SD, Johnston BC, Furukawa TA, Guyatt GH. Pooling health-related quality of life outcomes in meta-analysis: a tutorial and review of methods for enhancing interpretability. *Res Synth Methods* 2011;2:188-203. doi:10.1002/jrsm.46
- 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. doi:10.1016/j.jclinepi.2012.08.001
- Johnston BC, Patrick DL, Thorlund K, et al. Patient-reported outcomes in meta-analyses-part 2: methods for improving interpretability for decision-makers. *Health Qual Life Outcomes* 2013;11:211. doi:10.1186/1477-7525-11-211
- Mills K, Naylor JM, Roos EM, et al. Examining the minimal important difference of the knee injury and osteoarthritis outcome score. *Osteoarthritis Cartilage* 2015;23:A345-46. doi:10.1016/j.joca.2015.02.634
- Devji T, Guyatt GH, Lytvyn L, et al. Application of minimal important differences in degenerative knee disease outcomes: a systematic review and case study to inform *BMJ Rapid Recommendations*. *BMJ Open* 2017;7:e015587. doi:10.1136/bmjopen-2016-015587
- Higgins J, Li T, Deeks J. Chapter 6: choosing effect measures and computing estimates of effect. *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane, 2023:143-76.

- 37 Gallardo-Gomez D, Richardson R, Dwan K. Standardized mean differences in meta-analysis: A tutorial. *Cochrane Ev Synth* 2024;2:e12047. doi:10.1002/cesm.12047.
- 38 Hopkins WG, Rowlands DS. Standardization and other approaches to meta-analyze differences in means. *Stat Med* 2024;43:3092-108. doi:10.1002/sim.10114
- 39 Gallardo-Gómez D, Pedder H, Welton NJ, Dwan K, Dias S. Variability in meta-analysis estimates of continuous outcomes using different standardization and scale-specific re-expression methods. *J Clin Epidemiol* 2024;165:111213. doi:10.1016/j.jclinepi.2023.11.003
- 40 Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. Lawrence Erlbaum Associates, 1988.
- 41 Luo Y, Funada S, Yoshida K, Noma H, Sahker E, Furukawa TA. Large variation existed in standardized mean difference estimates using different calculation methods in clinical trials. *J Clin Epidemiol* 2022;149:89-97. doi:10.1016/j.jclinepi.2022.05.023
- 42 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
- 43 Devji T, Carrasco-Labra A, Qasim A, et al. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. *BMJ* 2020;369:m1714. doi:10.1136/bmj.m1714
- 44 Elwyn G, Durand MA, Agoritsas T, Mayer M. How to best convey continuous outcomes in patient decision aids. *BMJ Evid Based Med* 2024;bmjebm-2024-112871. doi:10.1136/bmjebm-2024-112871
- 45 Wiercioch W, Yepes-Núñez JJ, Schünemann H. Should any parenteral anticoagulation (UFH, LMWH or fondaparinux) vs. no parenteral anticoagulant be used in acutely ill medical patients for VTE prophylaxis? [Cited 26 Nov 2024.] <https://guidelines.gradeapro.org/profile/54B577E9-7F80-3A78-B3EA-3850E9A1D432>
- 46 Murad MH, Mustafa RA, Schünemann HJ, Sultan S, Santesso N. Rating the certainty in evidence in the absence of a single estimate of effect. *Evid Based Med* 2017;22:85-7. doi:10.1136/ebmed-2017-110668
- 47 McKenzie J, Brennan S. Synthesizing and presenting findings using other methods. In: Higgins J, Thomas J, Chandler J, et al, eds. *Cochrane handbook for Systematic Reviews of Interventions*. Version. 6.3 ed. Cochrane, 2022.
- 48 Tran VT, Barnes C, Montori VM, Falissard B, Ravaud P. Taxonomy of the burden of treatment: a multi-country web-based qualitative study of patients with chronic conditions. *BMC Med* 2015;13:115. doi:10.1186/s12916-015-0356-x
- 49 May CR, Eton DT, Boehmer K, et al. Rethinking the patient: using Burden of Treatment Theory to understand the changing dynamics of illness. *BMC Health Serv Res* 2014;14:281. doi:10.1186/1472-6963-14-281
- 50 Heen AF, Vandvik PO, Brandt L, et al. A framework for practical issues was developed to inform shared decision-making tools and clinical guidelines. *J Clin Epidemiol* 2021;129:104-13. doi:10.1016/j.jclinepi.2020.10.002
- 51 Agoritsas T, Heen AF, Brandt L, et al. Decision aids that really promote shared decision making: the pace quickens. *BMJ* 2015;350:g7624. doi:10.1136/bmj.g7624
- 52 Heen AF, Vandvik PO, Brandt L, et al. Decision aids linked to evidence summaries and clinical practice guidelines: results from user-testing in clinical encounters. *BMC Med Inform Decis Mak* 2021;21:202. doi:10.1186/s12911-021-01541-7
- 53 Cuello-Garcia CA, Santesso N, Morgan RL, et al. GRADE guidance 24 optimizing the integration of randomized and non-randomized studies of interventions in evidence syntheses and health guidelines. *J Clin Epidemiol* 2022;142:200-8. doi:10.1016/j.jclinepi.2021.11.026
- 54 WHO. Therapeutics and COVID-19. Living guideline in MAGICapp 2023. <https://app.magicapp.org/#/guideline/nBk01E>

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