

SERIES

GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions

Nancy Santesso^{a,*}, Claire Glenton^b, Philipp Dahm^c, Paul Garner^d, Elie A. Akl^e, Brian Alper^{f,g}, Romina Brignardello-Petersen^a, Alonso Carrasco-Labra^a, Hans De Beer^h, Monica Hultcrantzⁱ, Ton Kuijpers^j, Joerg Meerpohl^k, Rebecca Morgan^a, Reem Mustafa^{a,1}, Nicole Skoetz^m, Shahnaz Sultanⁿ, Charles Wiysonge^{o,p,q,r}, Gordon Guyatt^{a,s}, Holger J. Schünemann^{a,s}, for the GRADE Working Group

- ^aDepartment of Health Research Methods, Evidence and Impact, Cochrane Canada, MacGRADE Centre and Michael G. DeGroot Cochrane Canada Centre, McMaster University, 1280 Main St East, Hamilton, L8S 4L8, Canada
- ^bCochrane Norway and the Informed Health Choices Research Centre, Norwegian Institute of Public Health, Postboks 222 Skøyen, Sandakerveien 24C, inngang D11, 0213, Oslo, Norway
- ^cMinneapolis VA Health Care System, Urology Section 112D, One Veterans Drive, Minneapolis, MN, 55417, USA
- ^dCentre for Evidence Synthesis in Global Health, Liverpool School of Tropical Medicine, Liverpool, United Kingdom
- ^eDepartment of Internal Medicine, American University of Beirut, P.O.Box 11-0236, Lebanon
- ^fEBSCO Health, Innovations and Evidence-Based Medicine Development, 10 Estes Street, Ipswich, MA, 01938, USA
- ^gDepartment of Family and Community Medicine, University of Missouri-Columbia, Columbia, MO, USA
- ^hGuide2Guidance, Lemelerberg 7, 3524 LC Utrecht, the Netherlands
- ⁱSwedish Agency for Health Technology Assessment and Assessment of Social Services (SBU), S:t Eriksgatan 117, SE-102 33, Stockholm, Sweden
- ^jDepartment of Guideline Development and Research, Dutch College of General Practitioners (NHG), Mercatorlaan 1200, 3528, BL, Utrecht, the Netherlands
- ^kInstitute for Evidence in Medicine, Breisacher Strasse 153, 79110, Freiburg, Germany
- ^lDivision of Nephrology and Hypertension, Department of Internal Medicine, University of Kansas Medical Center, 3901 Rainbow Blvd, MS3002, Kansas City, KS, 66160, USA
- ^mFaculty of Medicine and University Hospital Cologne, Department I of Internal Medicine, University of Cologne, Kerpener Str. 62, 50931, Cologne, Germany
- ⁿDivision of Gastroenterology and Hepatology, and Nutrition, University of Minnesota, Minneapolis Veterans Affairs Healthcare System, 516 Delaware St. SE, 1st Floor, Phillips-Wangsteen Building, MMC 36, Minneapolis, MN, 55455, USA
- ^oCochrane South Africa, South African Medical Research Council, Cape Town, South Africa
- ^pSchool of Public Health and Family Medicine, University of Cape Town, Cape Town, South Africa
- ^qDepartment of Global Health, Stellenbosch University, Cape Town, South Africa
- ^rCochrane South Africa, South African Medical Research Council, Francie van Zijl Drive, Parow Valley, 7501, Cape Town, South Africa
- ^sDepartment of Medicine, McMaster University, 1280, Main St East, L8S 4L8, Hamilton, Canada

Accepted 7 October 2019; Published online 9 November 2019

Abstract

Objectives: Clear communication of systematic review findings will help readers and decision makers. We built on previous work to develop an approach that improves the clarity of statements to convey findings and that draws on Grading of Recommendations Assessment, Development and Evaluation (GRADE).

Study Design and Setting: We conducted workshops including 80 attendants and a survey of 110 producers and users of systematic reviews. We calculated acceptability of statements and revised the wording of those that were unacceptable to $\geq 40\%$ of participants.

Results: Most participants agreed statements should be based on size of effect and certainty of evidence. Statements for low, moderate and high certainty evidence were acceptable to $> 60\%$. Key guidance, for example, includes statements for high, moderate and low certainty for a large effect on intervention x as: x *results* in a large reduction...; x *likely results* in a large reduction...; x *may result* in a large reduction..., respectively.

Conclusions: Producers and users of systematic reviews found statements to communicate findings combining size and certainty of an effect acceptable. This article provides GRADE guidance and a wording template to formulate statements in systematic reviews and other

Declarations of interest: All authors confirm they have no direct financial conflicts of interest.

Funding: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

* Corresponding author. Department of Health Research Methods, Evidence and Impact, McMaster University, 1280 Main St East, Hamilton, L8S 4L8, Canada. Tel.: 1 289 407 1505; fax: 1 905 522 9507.
E-mail address: santesna@mcmaster.ca (N. Santesso).

<https://doi.org/10.1016/j.jclinepi.2019.10.014>

0895-4356/© 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

decision tools. © 2019 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Review literature as topic; Health communication; Evidence-based Medicine; Surveys and Questionnaires; Language; Persuasive communication

1. Introduction

Systematic reviews aim to synthesise evidence and provide readers with a summary of the findings for a specific intervention. To achieve this goal, the findings should be communicated as clearly and as simply as possible. The GRADE approach posits that there are two important components of a result of a review: the effect of the intervention, presented as the risk or difference in effect, as absolute numbers (e.g., 5 fewer deaths per 100), or as a narrative synthesis; and the certainty of (or confidence in) the evidence for that effect (categorised using the GRADE approach into high, moderate, low and very low) [1–6]. Both components should be conveyed to avoid misleading the reader. Consider, for example, a systematic review of the effects of waiving surgical fees to improve the use of cataract surgical services [7]. The authors found a risk ratio of 1.94 for the uptake of surgery, which they determined was an important increase in uptake. The certainty of evidence was low due to indirectness and imprecision (95% CI 1.14 to 3.31). If the authors conclude that there is an increase in uptake, but do not indicate that there is low certainty, readers could misinterpret the result as meaning that waiving surgical fees does increase uptake when in fact there is uncertainty. Although, the levels of evidence provided by the GRADE approach should be used to communicate the results (e.g., there is moderate certainty evidence that intervention A has X effect), various other phrases have been used, such as ‘limited evidence’, ‘insufficient evidence’, ‘no evidence to support’, or ‘the evidence shows, at best, a modest, non-statistically significant trend in favor of intervention A’. All of which can confuse readers. Previous research has explored methods to best communicate results and the GRADE Working Group has developed Evidence profiles and Summary of Findings Tables [3,8–10]. While these tables help readers understand the results of systematic reviews, this research found that many participants also appreciated brief statements describing the results [11,12].

However, guidance for how to interpret and communicate results using statements is limited. The previous version of the Cochrane Handbook provided some guidance to not describe results as statistically or not statistically significant and avoid the common misinterpretation that large *P* values mean ‘no difference’ or ‘no effect’ or small *P* values mean an important effect [1,8]. It also cautions authors about using ‘evidence of no effect’ or ‘no evidence of effect’ because these phrases are often used incorrectly. In 2010, we developed and tested four statements that were based on the size of an effect and the certainty of the evidence using the GRADE approach. Since

then, we have received informal feedback suggesting that these statements are restrictive and other options are needed, and therefore we decided to improve and test new approaches.

Our goal was to develop a set of standardized statements with multiple options for interpreting and communicating results of systematic reviews, and to write guidance. The statements assume that the evidence for an outcome is assessed using the GRADE approach or another formal system with four levels of evidence. It also assumes that certainty of evidence is not solely based on the imprecision of the result (i.e., power of the analysis and width of confidence interval), but also on other criteria, such as risk of bias of the studies, inconsistency (heterogeneity) of the result, indirectness (including subgroup analyses and applicability of the outcome measure), publication bias, and others.

2. Methods

2.1. Summary of research methods

The overall design is shown in [Figure 1](#).

2.2. Preliminary development

In 2010, during research to create a summary to present results from a systematic review to consumers, we developed, tested, and received feedback from an advisory group of statisticians about, statements to describe the effect of an intervention on an outcome. Single statements combined words for the size of an effect on an outcome and the certainty in that effect [12]. For example, suppose a review found that vitamin D results in an important reduction in falls with moderate certainty. The size of the effect would be described as *reduces*, and *probably* would indicate the certainty, and the final statement would be - “vitamin D probably reduces falls”. Depending on the size/importance of the effect, different qualifiers were used: for an important reduction in an outcome, the verb used was *reduces*; to describe a less important effect *slightly reduces* was used; and when the effect was close to a null effect, *little to no difference* was used. A different qualifier was used to express certainty: high, moderate, low or very low certainty were conveyed as *will*, *probably*, *may*, and *we are uncertain*, respectively.

During this research, we explored different approaches. Initially, we had six different ways to categorise the size of an effect based on how wide/narrow the confidence

What is new?

Key findings

- A set of statements to interpret results of systematic reviews of interventions and communicate them to patients, the public, and health care professionals was developed based on the GRADE approach to assess evidence. Experience with the statements and informal feedback showed that existing formulations were still not quite fit for purpose, and often used inconsistently.
- Building on results of workshops and a survey including producers and users of systematic reviews we revised the standardized statements.
- There was agreement that communicating the findings of reviews should be based on two components of a result: the magnitude or size of the effect and the certainty of the evidence.

What this adds to what was known

- Inconsistent words and phrases have been used to communicate the results of systematic reviews to users. Our suggested standardized statements are informative and were found to be acceptable to producers and users of systematic reviews. We provide detailed guidance for how to use the statements.

What is the implication and what should change now

- The template to formulate statements can be used to communicate the results of systematic reviews to users. These statements can be used in many sections of the systematic review, in evidence tables, and in tools or products for decision makers based on systematic reviews such as guideline recommendations.

intervals were. However, the width is already considered in the GRADE assessment and therefore the number of categories was reduced to three: *important*, *less important* and *little to no difference*. We also explored different qualifiers based on *why* evidence was rated down. If the evidence was low certainty because it was rated down twice for imprecision the qualifier was *we are very uncertain*, but if the evidence was rated down twice - once for imprecision and once for risk of bias - the qualifier was *possibly*. This system was after more discussion reduced to the four categories of GRADE because the level of certainty reflects our uncertainty regardless of what specific domains are rated down.

2.3. Workshops

Following publication of the minimum set of statements and years of informal feedback, a small working group of authors met and created a longer list of options. We conducted three workshops at GRADE meetings in 2016 and 2017, each with approximately 20 – 40 people with expertise in methods of systematic reviews and guideline development, some of whom did not speak English as a first language. During the workshops, participants reviewed 4–6 examples of the results for an outcome of a systematic review as forest plot of a meta-analysis (Figure 2), a narrative synthesis, or in absolute effects, along with the certainty of the evidence and explanations. We asked participants to discuss what statements they would use to express the result or if they agreed with the statement provided and why. We used the feedback to make revisions to our list.

2.4. Survey

From March to April 2018, we conducted an electronic survey using SurveyMonkey to determine the acceptability of the statements (Appendix 1). We purposively invited by email: 1) people who conduct or summarise systematic reviews for use in decision making; 2) people who use systematic reviews; and 3) statisticians with systematic review experience. Members of the GRADE Working Group were also invited. Invited participants could forward the email to others and we sent one reminder 1 week later. The survey link was also sent via one author's professional Twitter account (approximately 2000 followers). The first part of the survey asked participants about their roles in reviews and epidemiological training. Section 2 presented results for one outcome from five systematic reviews with 3 to 4 statements. Respondents rated the statements as unacceptable, acceptable or ideal. Section 3 asked 'Do you agree in principle that conclusions should be based on the concepts of the importance/size of the effect and the certainty of the evidence?'. We piloted the survey in two people and revised accordingly. The Hamilton Integrated Research Ethics Board waived formal ethics approval. One investigator analysed the data using descriptive statistics, and summarised the free-text comments by broad themes. A priori, we decided to revise statements that were 'unacceptable' to more than 40% and keep statements that more than 60% judged acceptable or ideal.

2.5. Incorporation of results

The lead authors incorporated the survey and workshop results into the statements and developed guidance. We presented the results to approximately 60 attendees at a GRADE Working Group meeting (April 2018) and to approximately 80 people in September 2018 (for approval).

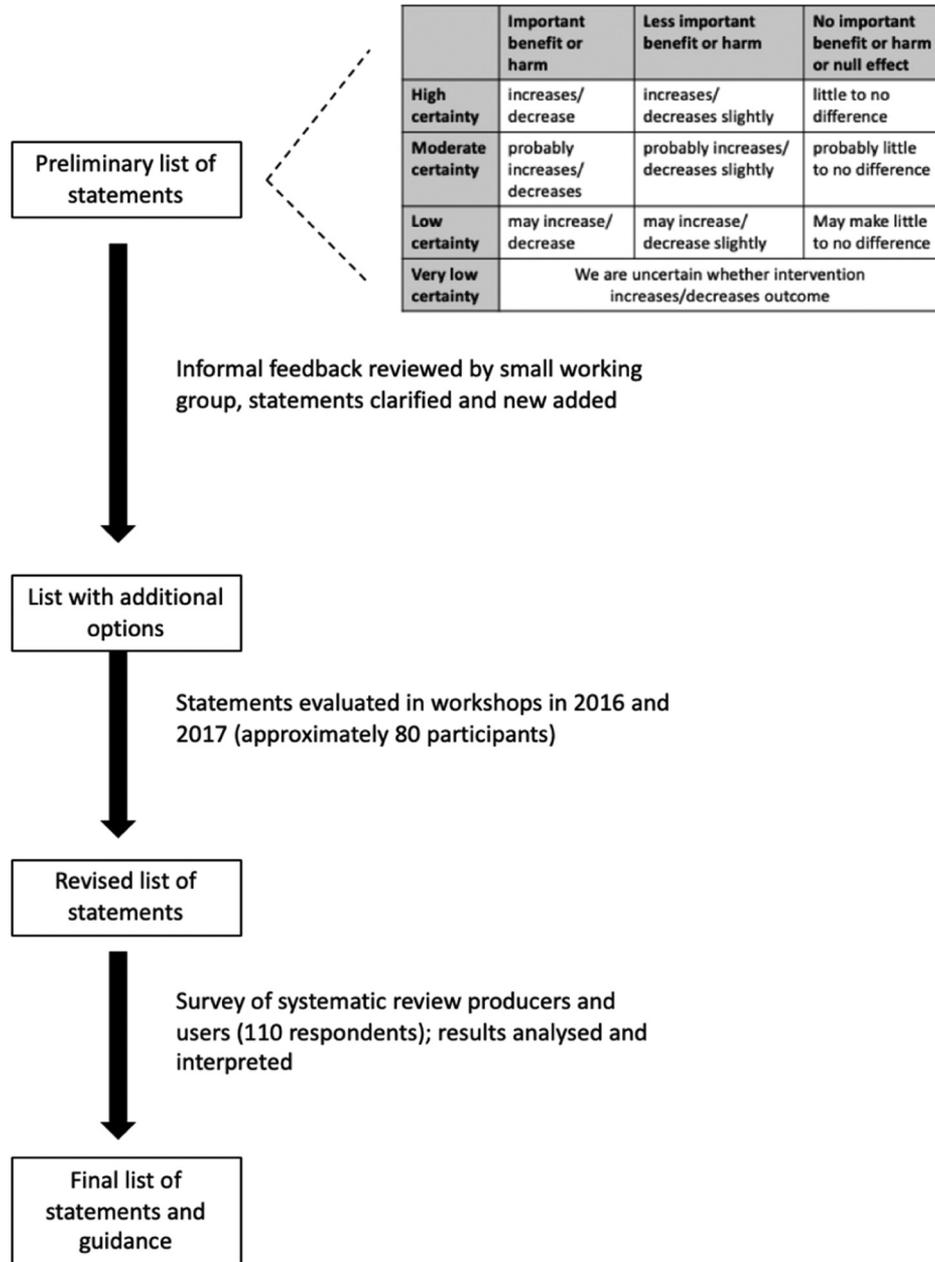


Fig. 1. Study design.

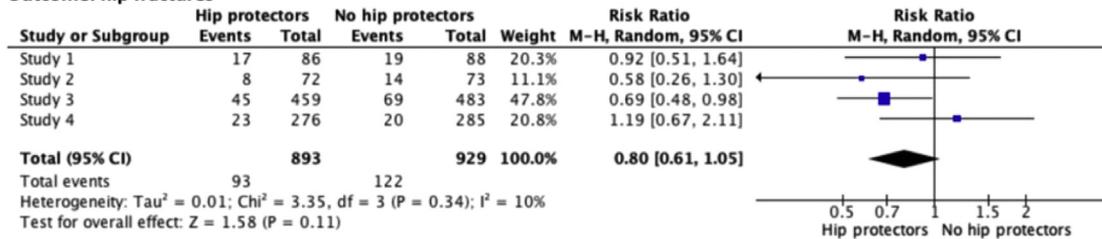
3. Results and implications

3.1. Acceptability of statements

Of the 110 respondents (19 of whom were members of this GRADE project group), 72% described themselves as systematic review or guideline methodologists, and 13% as readers of reviews. Approximately, 30% indicated they had no formal education in epidemiology. Two did not answer all questions; however, their results were included. In section 2, 39 provided written comments about acceptability, and 15 provided comments in section 3. We present results from the 91 participants and use the comments of

the project members to contextualise results (see [Appendix 2](#) for raw data from survey). We did not calculate a response rate since participants could forward the link to others. The final list of informative statements is in [Table 1](#).

Acceptability of statements for very low certainty evidence: The statement “[Intervention X] may reduce the [outcome] slightly but we are uncertain” was presented in two examples and was rated as unacceptable by 37% in one example and 46% in the other. The comments highlighted that *we are uncertain* could be misinterpreted; respondents suggested that it would be clearer to instead write that *the evidence* is uncertain. The two examples also provided two statements stating

Outcome: hip fractures

A review of hip protectors versus no hip protectors to prevent hip fractures at 1 year found risk ratio (RR) 0.80 (95% CI, 0.61 to 1.05). This means in 100 people at high risk of hip fractures (10 per 100) that there would be 2 fewer fractures (from 4 fewer to 1 more). The evidence was assessed as moderate certainty because of lack of blinding in the studies and some imprecision due to few events. The authors indicated that any reduction in fractures is beneficial but 5 fewer is important.

The abstract included the statement “hip protectors probably reduces the risk of hip fractures slightly”. Do you agree with this statement? Why or why not?

Fig. 2. Example of information provided to workshop participants for feedback. Note: the appropriate statement in this example is ‘hip protectors probably reduces the risk of hip fractures slightly’.

the direction of effect: “We are uncertain about whether co-enzyme Q10 reduces blood pressure” – acceptable to 80%, and “We are uncertain about the effect of co-enzyme Q10 on blood pressure” – acceptable to 71%. During workshops, there was also some debate about communicating a direction of effect when the evidence is so uncertain. However, we have kept both options for very low certainty: uncertain effect with or without a direction of effect.

Acceptability of statements for low certainty evidence: Participants were presented with the qualifying words *may*, *appears*, *suggests*, and *likely* (“Probiotics may result in a large reduction in the incidence of diarrhea). *Likely* was rated as unacceptable by 52%; *appears* by 50%, and *suggests* by 57%. Respondents observed that most words to convey low certainty evidence were vague e.g., *may* could be interpreted may or may not. Respondents wrote that *suggests* could be more acceptable, and some noted that *appears* sounded supernatural. Therefore, *appears* was deleted, but *may* and *suggests* remain options for low certainty evidence.

Acceptability of statements for moderate certainty and high certainty evidence: There were few comments and both *likely* and *probably* were acceptable.

Acceptability of statements to communicate size of effect: In one example, the intervention resulted in 2 more hip fractures per 1000 (from 2 fewer to 6 more) and the authors judged that 2 more did not reach a threshold for an effect either as a beneficial reduction or as a harm. Two of the example narrative statements used *results in little to no difference* and the other two used *does not reduce outcome*. *Little to no difference* was unacceptable to 20%, and *does not reduce* to 35–40%. There were many comments that *does not* should not be used when communicating a result close to null effect. Workshop participants also often expressed concern with interpreting null effect as *does not* affect.

Another example explored the acceptability of statements to convey evidence for a small effect that is not important. Two of the three statements describing the effect as a *small possible unimportant reduction* were rated as unacceptable by 45% to 50%. Participants responded that the high number of qualifying words could be confusing. Statements with multiple qualifiers for importance were therefore deleted and a *small effect* has been divided into a *small and important effect* and an unimportant effect as *trivial or small, unimportant or no effect* (*trivial* is added to be consistent with GRADE’s Evidence to Decision frameworks [13–17]). In this example, *do not result in* was used and again there were comments that it is not correct to describe a result near the null effect as *not* occurring. The words *do not* or *does not* to describe *little to no effect* are still an option.

3.2. Agreement about principles of size of effect and certainty of evidence

Ninety-nine percent (84/85) agreed that statements should be based on both size of the effect and certainty of evidence. In general, respondents were concerned that it is difficult to determine whether an effect is large, moderate, small (important or not important), or of little to no effect. Comments also highlighted to not interpret wide confidence intervals and non-statistically significant results as *no effect*.

4. Discussion and guidance

4.1. Discussion

We have created a list of brief and informative statements that authors of systematic reviews, and people

Table 1. Final list of informative statements to communicate results of systematic reviews

Size of the effect estimate	Suggested statements (replace X with intervention, replace 'reduce/increase' with direction of effect, replace 'outcome' with name of outcome, include 'when compared with Y' when needed)
HIGH Certainty of the evidence	
Large effect	X results in a large reduction/increase in outcome
Moderate effect	X reduces/increases outcome X results in a reduction/increase in outcome
Small important effect	X reduces/increases outcome slightly X results in a slight reduction/increase in outcome
Trivial, small unimportant effect or no effect	X results in little to no difference in outcome X does not reduce/increase outcome
MODERATE Certainty of the evidence	
Large effect	X likely results in a large reduction/increase in outcome X probably results in a large reduction/increase in outcome
Moderate effect	X likely reduces/increases outcome X probably reduces/increases outcome X likely results in a reduction/increase in outcome X probably results in a reduction/increase in outcome
Small important effect	X probably reduces/increases outcome slightly X likely reduces/increases outcome slightly X probably results in a slight reduction/increase in outcome X likely results in a slight reduction/increase in outcome
Trivial, small unimportant effect or no effect	X likely results in little to no difference in outcome X probably results in little to no difference in outcome X likely does not reduce/increase outcome X probably does not reduce/increase outcome
LOW Certainty of the evidence	
Large effect	X may result in a large reduction/increase in outcome The evidence suggests X results in a large reduction/increase in outcome
Moderate effect	X may reduce/increase outcome The evidence suggests X reduces/increases outcome X may result in a reduction/increase in outcome The evidence suggests X results in a reduction/increase in outcome
Small important effect	X may reduce/increase outcome slightly The evidence suggests X reduces/increases outcome slightly X may result in a slight reduction/increase in outcome The evidence suggests X results in a slight reduction/increase in outcome
Trivial, small unimportant effect or no effect	X may result in little to no difference in outcome The evidence suggests that X results in little to no difference in outcome X may not reduce/increase outcome The evidence suggests that X does not reduce/increase outcome
VERY LOW Certainty of the evidence	
Any effect	The evidence is very uncertain about the effect of X on outcome X may reduce/increase/have little to no effect on outcome but the evidence is very uncertain

presenting evidence to decision makers, e.g., guideline developers, can use to describe the results (Table 1). This work builds on our previous research, on many years of experience using the statements, a survey, and on feedback received during GRADE working group meetings. Although we piloted examples and the survey, there is still the potential that we may not have expressed the task clearly to respondents, resulting in some confusion. However, we received comments from a variety of important stakeholders, including methodologists in systematic reviews and guidelines and readers, and found results were

consistent. We provide guidance to use these statements, and examples in Appendix 3.

4.2. Use of certainty of evidence and size of effect to write informative statements

The basic premise is that review authors should report both the effect of an intervention on an outcome and the certainty in the evidence. Authors can communicate these components in multiple ways. GRADE guidance now suggests two approaches. First, authors may communicate the

findings by providing the effect on the outcome and the certainty of the evidence according to the GRADE levels of evidence (i.e., provide the point estimate and confidence interval in relative and absolute terms, and then specify that the evidence is “moderate certainty”). Second, if authors want to communicate the result in one statement, they should use [Table 1](#), first selecting the category for certainty of evidence, then making a judgment regarding the size of the effect, and finally choosing from the appropriate wording options (e.g., for a small important effect of moderate certainty - “intervention A likely increases outcome X slightly”).”

4.3. Decisions about the size of the effect

To create a statement using [Table 1](#), authors must decide into which category the size of effect falls. The GRADE Evidence to Decision framework provides some guidance about the size of effect [13–17]. However, when conducting a GRADE assessment, in particular when assessing imprecision, systematic reviewers partially contextualise decisions using thresholds for *no or trivial, small, moderate and large effects* [18–21]. These decisions can be based on research into minimal important differences, discussions within the systematic review team, or consultation with decision-makers, and should be transparent. Two considerations are of critical importance when determining the size. The first is calculating and using absolute effects rather than using relative effects that can often be misleading. For instance, consider a risk ratio 0.84, or 16% relative reduction in hip fractures in older adults. If on the one hand, the baseline risk of hip fractures is 20/1000 over 1 year, the risk ratio 0.84 would translate to 3 fewer per 1000, which most would consider a small effect. On the other hand, if the baseline risk is 200/1000, many would consider that the resulting absolute reduction of 32 per 1000 is a moderate to large effect. The second is identifying the value of the outcome [16,17]. Ideally, review authors identify the thresholds, and use them to rate the certainty of the evidence. The approach to choose a threshold (or range) can be either fully contextualised (based on consideration of all critical outcomes) or partially contextualised (based on the value of the individual outcome.) [20]. Whatever the thresholds, a decision needs to be made in order to write a statement using [Table 1](#).

When deciding on thresholds, review authors also need to be aware of the risk of misinterpreting a result with a wide confidence interval that includes ‘1’ (for relative effects) or ‘0’ (for absolute effects) as ‘no effect’ or ‘no difference’ [22,23]. For example, consider a mean difference for the effect of a treatment on quality of life is 1.5 (95% CI, –1.2 to 4.2) where an important effect is an increase of 1 on a scale of 1 to 10 (better), and the certainty of the evidence is low (due to imprecision and risk of bias). The point estimate is an increase of 1.5, and we would characterise the effect as important,

Box 1 Best estimate vs. confidence intervals to determine effect size

The statements communicate the size of the effect based on the point estimate in a meta-analysis or on the summary estimate in a narrative synthesis instead of the confidence intervals. Confidence intervals represent the range in which a point estimate would fall if multiple experiments were conducted, or as the range of values either side of the estimate between which we can be 95% sure that the true value lies [22], and are calculated based on factors such as sample sizes and variance within or between studies. The calculation does not factor in the risk of bias of the studies; indirectness of the populations, interventions or outcomes; or, the risk of publication bias, which (if there were methods to do so) could widen the confidence intervals, making the calculated confidence intervals meaningless. However, when conducting a GRADE assessment authors consider the width of the confidence intervals and power of the analysis (i.e., imprecision) plus all of the other factors to determine the certainty of the evidence. Thus, the certainty around the point estimate varies depending on what domains demonstrate shortcomings and except for imprecision that certainty interval is not known [18,19]. For this reason, when communicating an effect using statements, authors should focus on the best estimate and on the certainty in that estimate which considers multiple factors.

likely moderate, but not ‘no effect’. Authors need to determine the size of the effect based on the effect estimate, not on the confidence intervals. The width of the confidence interval is considered in the assessment of the certainty of the evidence (see [Box 1](#)). In this case, the certainty is low, we use the word ‘may’, and the final statement is, ‘the [treatment] may increase quality of life’. In contrast, if the effect was an increase of 0.3 (95% CI, –1.8 to 2.3), the effect could be categorised as ‘trivial, small unimportant or no effect’ because the effect estimate is less than our threshold for an important difference, and the final statement based on low certainty evidence would be ‘the [treatment] may have little to no effect on quality of life’.

4.4. Use the statements in the text of a review and in summary of findings tables

Authors can use these statements throughout a systematic review: in the abstract, plain language summary,

The screenshot displays the GRADEpro interface for a clinical question: "Should a ketogenic diet for 3-6 months vs. non-ketogenic diet be used in people with type II diabetes?". The main table lists outcomes such as HbA1c levels, fasting blood glucose, BMI, weight, waist circumference, and quality of life. An "Automatic narrative" dialog box is open, showing a dropdown menu for "Size of effect" with options: None, Large effect, Moderate effect, Small effect (important), Small effect (not important), and No effect. The "Narrative statements" field contains the text: "A ketogenic diet for 3-6 months probably reduces ch...".

Outcome No of participants (studies)	Relative effect (95% CI)	Anticipated absolute effects (95% CI)		Difference	Certainty	What happens
		Without a ketogenic diet for 3-6 months	With a ketogenic diet for 3-6 months			
Change in HbA1c levels No of participants: 592 (7 RCTs)	-	The mean change in HbA1c levels without a ketogenic diet for 3-6 months was 1 %	-			
Change in fasting blood glucose assessed with: mg/dL No of participants: 462 (5 RCTs)	-	The mean change in fasting blood glucose without a ketogenic diet for 3-6 months was 20 mg/dL	-			
Body Mass Index (BMI) No of participants: 266 (3 RCTs)	-	The mean body Mass Index without a ketogenic diet for 3-6 months was 32	-			
Change in weight assessed with: kg No of participants: 558 (7 RCTs)	-	The mean change in weight without a ketogenic diet for 3-6 months was 5 kg	-			
Waist circumference assessed with: cm No of participants: 262 (3 RCTs)	-	The mean waist circumference without a ketogenic diet for 3-6 months was 100 cm	-			
Change in Quality of life	-	The mean change in Quality of	-			

Fig. 3. Screenshot of GRADEpro and automatic generation of informative statements based on size of effect and certainty of evidence.

results, discussion, and in evidence tables. Experience has shown that this approach to wording should not be an automated application, which could result in a list of monotonous statements. In GRADEpro (www.grade-pro.org), the software programme to produce summary of findings tables, the size of effect and the certainty of evidence are used to automatically generate an editable statement (Figure 3).

Systematic reviews typically compare an intervention/test to a comparator. The statements in Table 1 do not explicitly state the comparator which may be acceptable when the comparator is standard care, a placebo, or no intervention, but when it is an alternative intervention, it's important to include it. Using a hypothetical example, there is low certainty evidence that oseltamivir reduced the duration of symptoms by 2 days (95% CI, 0.5 to 3.6 days) when compared to zanamivir, whereby 2 days was an important difference. The informative statement should be 'oseltamivir may reduce the duration of symptoms more than zanamivir'.

4.5. Borderline decisions and very low certainty of the evidence

When applying the GRADE approach, authors may debate about the weight of each domain to determine the level of evidence. For example, in some cases, moderate certainty evidence may be due solely to imprecision, in other cases, it may be a combination of small concerns with imprecision, risk of bias and inconsistency. Despite these differences, authors must make a final decision about the level of evidence, and it is this level that determines the wording options available to use in that category. The GRADE approach to certainty of evidence, however, acknowledges that, despite the four categories of high, moderate, low and

very low, certainty is a continuum [2]. Consequently, users may find that when deciding on the certainty they may have been on the threshold between categories, but ultimately had to choose a category, make a borderline decision, or characterise the certainty as being at a threshold. When choosing a statement in these instances, users could choose from the statements on either side of the border.

We have also provided two options for a statement based on very low certainty of evidence: one option gives the direction of the effect, the other does not. Ratings are on a continuum and within the category of very low there may be situations when authors feel somewhat more compelled to express an effect (e.g., when the rating borders on low) and situations when they do not (e.g., the evidence is at the very bottom of the continuum of certainty).

4.6. Use of the statements in different review types

The underlying principle considering size of effect and certainty of evidence (whether GRADE or another system with four levels) to write statements can likely be applied to any review type. In a test accuracy review with pooled sensitivity and specificity estimates, the absolute numbers of misidentified people (i.e., false negatives and positives) can be quantified as large, moderate, small, or trivial, depending on the consequences for patients. A review may find that a cytology test misses 20 more out of 1000 women with cervical cancer lesions than an HPV test - a small difference based on moderate certainty evidence. We could conclude that 'when compared to HPV tests, cytology tests *probably* miss *slightly* more women with cervical lesions.' In prognostic reviews, the statements could be written as 'associations'. For example, for a moderately sized association of hip fractures with age and low certainty evidence, the statement would be 'age *may* be associated with hip fractures'.

5. Conclusions

The informative statements to communicate results of systematic reviews should be used throughout the text of a systematic review, in the abstract, plain language summary, results, discussion, and in evidence tables. These statements can also be used in other tools and products that communicate the results of systematic reviews to decision makers, and in fact are already being used in health care guidelines to summarise the evidence and in patient versions of guidelines [24–26]. The list was also originally translated into Spanish, Norwegian, Italian, French and German [12], and future work will focus on these translations.

CRedit authorship contribution statement

Nancy Santesso: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Writing - original draft, Writing - review & editing. **Claire Glenton:** Conceptualization, Writing - review & editing. **Philipp Dahm:** Conceptualization, Writing - review & editing. **Paul Garner:** Conceptualization, Writing - review & editing. **Elie A. Akl:** Conceptualization, Writing - review & editing. **Brian Alper:** Conceptualization, Writing - review & editing. **Romina Brignardello-Petersen:** Conceptualization, Writing - review & editing. **Alonso Carrasco-Labra:** Conceptualization, Writing - review & editing. **Hans De Beer:** Conceptualization, Writing - review & editing. **Monica Hulcrantz:** Conceptualization, Writing - review & editing. **Ton Kuijpers:** Conceptualization, Writing - review & editing. **Joerg Meerpohl:** Conceptualization, Writing - review & editing. **Rebecca Morgan:** Conceptualization, Writing - review & editing. **Reem Mustafa:** Conceptualization, Writing - review & editing. **Nicole Skoetz:** Conceptualization, Writing - review & editing. **Shahnaz Sultan:** Conceptualization, Writing - review & editing. **Charles Wiysonge:** Conceptualization, Writing - review & editing. **Gordon Guyatt:** Conceptualization, Methodology, Writing - review & editing. **Holger J. Schünemann:** Conceptualization, Methodology, Writing - review & editing.

Acknowledgments

We would also like to acknowledge specific GRADE Working Group members that provided help with the project: Arnav Agarwal, Sarah Rosenbaum, Jasvinder Singh, Airton Stein, Judith Thornton, Gemma Villanueva, and Lee Yee Chong.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2019.10.014>.

References

- [1] Schünemann HJOA, Vist GE, Higgins JPT, Deeks JJ, Glasziou P, Guyatt GH. Chapter 12: interpreting results and drawing conclusions. In: Higgins JPTGS, editor. *Cochrane Handbook for Systematic Reviews of Interventions* Version 510 (updated March 2011). The Cochrane Collaboration; 2008:2008. Available from www.cochrane-handbook.org.
- [2] Guyatt G, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, et al. GRADE guidelines 11-making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. *J Clin Epidemiol* 2013;66:151–7.
- [3] Guyatt G, Oxman AD, Akl EA, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64:383–94.
- [4] Guyatt GH, Oxman AD, Schunemann HJ, Tugwell P, Knottnerus A. GRADE guidelines: a new series of articles in the *Journal of Clinical Epidemiology*. *J Clin Epidemiol* 2010;64:380–2.
- [5] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;336:924–6.
- [6] Schunemann HJ, Best D, Vist G, Oxman AD. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. *CMAJ* 2003;169(7):677–80.
- [7] Ramke J, Petkovic J, Welch V, Bagnault I, Gilbert C, Blanchet K, et al. Interventions to improve access to cataract surgical services and their impact on equity in low- and middle-income countries. *Cochrane Database Syst Rev* 2017;11:CD011307.
- [8] Schünemann HJHJ, Vist GE, Glasziou P, Akl E, Skoetz N, Guyatt GH. Chapter 14: completing summary of findings tables and grading the certainty of evidence. In: Higgins JPT, Thomas J, Chandler J, Cumston M, Li T, PageMJ Welch V, editors. *Cochrane Handbook for Systematic Reviews of Interventions* Version 6 (updated January 29, 2019). The Cochrane Collaboration; 2019:2019. Available from <https://training.cochrane.org/handbooks>.
- [9] Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, et al. GRADE guidelines 12. Preparing summary of findings tables-binary outcomes. *J Clin Epidemiol* 2013;66:158–72.
- [10] Schünemann HJOA, Higgins JPT, Vist GE, Glasziou P, Guyatt GH. Chapter 11: presenting results and ‘Summary of findings’ tables. In: Higgins JPTGS, editor. *Cochrane Handbook for Systematic Reviews of Interventions* Version 510 (updated March 2011). The Cochrane Collaboration; 2008:2008. Available from www.cochrane-handbook.org.
- [11] Carrasco-Labra A, Brignardello-Petersen R, Santesso N, Neumann I, Mustafa RA, Mbuagbaw L, et al. Improving GRADE evidence tables part 1: a randomized trial shows improved understanding of content in summary of findings tables with a new format. *J Clin Epidemiol* 2016;74:7–18.
- [12] Glenton C, Santesso N, Rosenbaum S, Nilsen ES, Rader T, Ciapponi A, et al. Presenting the results of Cochrane Systematic Reviews to a consumer audience: a qualitative study. *Med Decis Making* 2010;30:566–77.
- [13] Moberg J, Oxman AD, Rosenbaum S, Schunemann HJ, Guyatt G, Flottorp S, et al. The GRADE Evidence to Decision (EtD) framework for health system and public health decisions. *Health Res Policy Syst* 2018;16(1):45.
- [14] Schunemann HJ, Wiercioch W, Brozek J, Etzeandia-Ikobaltzeta I, Mustafa RA, Manja V, et al. GRADE Evidence to Decision (EtD) frameworks for adoption, adaptation, and de novo development of trustworthy recommendations: GRADE-ADOLOPMENT. *J Clin Epidemiol* 2017;81:101–10.
- [15] Parmelli E, Amato L, Oxman AD, Alonso-Coello P, Brunetti M, Moberg J, et al. GRADE EVIDENCE TO DECISION (EtD) FRAMEWORK FOR COVERAGE DECISIONS. *Int J Technol Assess Health Care* 2017:1–7.
- [16] Alonso-Coello P, Schunemann HJ, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD)

- frameworks: a systematic and transparent approach to making well informed healthcare choices. 1: introduction. *BMJ* 2016;353:i2016.
- [17] Alonso-Coello P, Oxman AD, Moberg J, Brignardello-Petersen R, Akl EA, Davoli M, et al. GRADE Evidence to Decision (EtD) frameworks: a systematic and transparent approach to making well informed healthcare choices. 2: clinical practice guidelines. *BMJ* 2016;353:i2089.
- [18] Schunemann HJ. Interpreting GRADE's levels of certainty or quality of the evidence: GRADE for statisticians, considering review information size or less emphasis on imprecision? *J Clin Epidemiol* 2016;75:6–15.
- [19] Anttila S, Persson J, Varemán N, Sahlin NE. Conclusiveness resolves the conflict between quality of evidence and imprecision in GRADE. *J Clin Epidemiol* 2016;75:1–5.
- [20] Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. *J Clin Epidemiol* 2017;87:4–13.
- [21] Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence—imprecision. *J Clin Epidemiol* 2011;64:1283–93.
- [22] Altman DG. Why we need confidence intervals. *World J Surg* 2005;29:554–6.
- [23] Nuzzo R. Scientific method: statistical errors. *Nature* 2014;506:150–2.
- [24] Papaioannou A, Santesso N, Morin SN, Feldman S, Adachi JD, Crilly R, et al. Recommendations for preventing fracture in long-term care. *CMAJ* 2015;187(15):1135–44. e1450-e1161.
- [25] Santesso N, Carrasco-Labra A, Brignardello-Petersen R. Hip protectors for preventing hip fractures in older people. *Cochrane Database Syst Rev* 2014Cd001255.
- [26] Wieland LS, Santesso N. A summary of a Cochrane review: Acupuncture or acupressure for induction of labour. *Eur J Integr Med* 2018;17:141–2.