

# Improving GRADE evidence tables part 3: detailed guidance for explanatory footnotes supports creating and understanding GRADE certainty in the evidence judgments

Nancy Santesso<sup>a,b,c</sup>, Alonso Carrasco-Labra<sup>a,b,c,d</sup>, Miranda Langendam<sup>b,e</sup>,  
Romina Brignardello-Petersen<sup>f,g,h</sup>, Reem A. Mustafa<sup>a,b,h,i</sup>, Pauline Heus<sup>j</sup>, Toby Lasserson<sup>k</sup>,  
Newton Opiyo<sup>k</sup>, Ilkka Kunnamo<sup>l</sup>, David Sinclair<sup>m</sup>, Paul Garner<sup>m,n</sup>, Shaun Treweek<sup>o</sup>,  
David Tovey<sup>k</sup>, Elie A. Akl<sup>a,b,p</sup>, Peter Tugwell<sup>q</sup>, Jan L. Brozek<sup>a,b,c,r</sup>, Gordon Guyatt<sup>a,b,c,r</sup>,  
Holger J. Schünemann<sup>a,b,c,r,\*</sup>

<sup>a</sup>Department of Clinical Epidemiology & Biostatistics, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada

<sup>b</sup>Cochrane GRADEing Methods Group, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada

<sup>c</sup>McMaster GRADE Center, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada

<sup>d</sup>Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, Universidad de Chile, Sergio Livingstone Pohlhammer 943, Independencia, Santiago, Chile

<sup>e</sup>Department of Clinical Epidemiology, Biostatistics and Bioinformatics, Academic Medical Center, University of Amsterdam, PO Box 22660, J1B-211, 1100 DD Amsterdam, The Netherlands

<sup>f</sup>Evidence-Based Dentistry Unit, Faculty of Dentistry, Universidad de Chile, Sergio Livingstone Pohlhammer 943, Independencia, Santiago, Chile

<sup>g</sup>Institute of Health Policy, Management and Evaluation, University of Toronto, 155 College Street, 4th Floor, Toronto, Ontario M5T 3M6, Canada

<sup>h</sup>Department of Medicine/Nephrology, University of Missouri-Kansas City, 2411 Holmes Street, Kansas City, MO 64108-2792, USA

<sup>i</sup>Department of Biomedical & Health Informatics, University of Missouri-Kansas City, 2411 Holmes Street, Kansas City, MO 64108-2792, USA

<sup>j</sup>Cochrane Netherlands, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Huispostnummer Street 6.131, Postbus 85500, 3508 GA Utrecht, The Netherlands

<sup>k</sup>Cochrane Editorial Unit, St Albans House, 57-59 Haymarket, London SW1Y 4QX, UK

<sup>l</sup>Department of General Practice and Primary Health Care, Duodecim Medical Publications Ltd and University of Helsinki, PO Box 874, Kaivokatu 10 A, 7th Floor, FIN-00101 Helsinki, Finland

<sup>m</sup>Department of Clinical Sciences, Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK

<sup>n</sup>Cochrane Infectious Diseases Group, Pembroke Place, Liverpool L3 5QA, UK

<sup>o</sup>Health Services Research Unit, University of Aberdeen, Health Sciences Building, Foresterhill, Aberdeen AB25 2ZD, UK

<sup>p</sup>Department of Internal Medicine, American University of Beirut, Riad-El-Solh, PO Box 11-0236, Beirut 1107 2020, Lebanon

<sup>q</sup>Clinical Epidemiology Unit, Ottawa Hospital Research Institute, Ottawa Hospital, Ottawa, Ontario, Canada

<sup>r</sup>Department of Medicine, McMaster University, 1280 Main Street West, Hamilton, Ontario L8S 4K1, Canada

Accepted 22 December 2015; Published online 12 January 2016

## Abstract

**Background:** The Grading of Recommendations Assessment, Development and Evaluation (GRADE) is widely used and reliable and accurate for assessing the certainty in the body of health evidence. The GRADE working group has provided detailed guidance for assessing the certainty in the body of evidence in systematic reviews and health technology assessments (HTAs) and how to grade the strength of health recommendations. However, there is limited advice regarding how to maximize transparency of these judgments, in particular through explanatory footnotes or explanations in Summary of Findings tables and Evidence Profiles (GRADE evidence tables).

**Methods:** We conducted this study to define the essential attributes of useful explanations and to develop specific guidance for explanations associated with GRADE evidence tables. We used a sample of explanations according to their complexity, type of judgment involved, and appropriateness from a database of published GRADE evidence tables in Cochrane reviews and World Health Organization guidelines. We used an iterative process and group consensus to determine the attributes and develop guidance.

**Results:** Explanations in GRADE evidence tables should be concise, informative, relevant, easy to understand, and accurate. We provide general and domain-specific guidance to assist authors with achieving these desirable attributes in their explanations associated with GRADE evidence tables.

Conflict of interest: The authors are members of the GRADE working group and/or the Cochrane Collaboration. None of them declared a direct financial conflict of interest.

\* Corresponding author. Tel.: 9055259140; fax: 9055229507.  
E-mail address: schuneh@mcmaster.ca (H.J. Schünemann).

**Conclusions:** Adhering to the general and GRADE domain-specific guidance should improve the quality of explanations associated with GRADE evidence tables, assist authors of systematic reviews, HTA reports, or guidelines with information that they can use in other parts of their evidence synthesis. This guidance will also support editorial evaluation of evidence syntheses using GRADE and provide a minimum quality standard of judgments across tables. © 2016 Elsevier Inc. All rights reserved.

*Keywords:* Levels of evidence; GRADE; Summary of Findings table; Evidence Profile; Explanations; Guidelines; GRADEpro; Systematic reviews

**1. Introduction**

The Grading of Recommendations, Assessment, Development and Evaluation (GRADE) approach was developed over 15 years ago with the purpose of providing a common, standardized, explicit, and transparent method for evaluating the certainty in the evidence and formulating health recommendations [1,2]. Whether this approach is used in the context of systematic reviews, health technology assessment (HTA) reports, or guidelines, its optimal application requires adhering to appropriate methods [3].

GRADE suggests to present syntheses of evidence using “GRADE evidence tables”: detailed Evidence Profiles (EPs) and/or Summary of Findings (SoFs) tables (see Box 1) [3]. They are considered the critical source of information of any evidence synthesis. EPs are intended to provide detailed judgments regarding each domain that determines the certainty in the evidence, also known as quality of the evidence, strength of evidence, or confidence in effect estimates. The main audiences for these profiles are audiences who have an interest in detailed understanding of the judgments determining the certainty in the evidence, for example, guideline panels. Thus, EPs are required to fully explain judgments. They are necessary to develop an SoF table that represents an alternative, more compact presentation format. SoF tables are typically intended for a broader audience, including, for example, health professionals and policy makers.

*1.1. Importance of explanations to ensure transparency*

Although the application of GRADE is reliable and accurately reflects the certainty in the body of evidence, current practice for developing GRADE evidence tables varies [4]. Because judgments are required and may differ based on context, this variability in judgments is expected. Variability across raters may not be problematic, as long as effective explanations for judgments are provided. Providing the rationale for each decision allows users to either agree or disagree with the decisions. GRADE requires producers of evidence tables to provide explanatory footnotes, which GRADE now also calls “explanations.”

These explanations represent the opportunity for developers of such tables to explain, for example, the rationale for downgrading or upgrading the certainty in the evidence, the source for baseline risks, or other information that facilitates the interpretation of the data displayed. The use of these explanations provides the GRADE approach with transparency, one of its most important features.

Thus far, the GRADE working group has provided detailed and explicit guidance about how to assess the certainty in the evidence and grading the strength of recommendations but offered no detailed advice regarding how to maximize transparency of these judgments through the use of explanations [3,5–9]. This article is the final in a series of three articles that focus on improving GRADE evidence tables, which was in part supported by the Cochrane Methods Innovation Fund. The first article reported the

**Box 1 Content of Evidence profiles and Summary of findings tables**

<b>Evidence profile</b>	<b>Summary of findings table</b>
More detailed summary of findings	Compact summary of findings
Patient important outcomes	Patient important outcomes
Relative and absolute effect estimates	Relative and absolute effect estimates
Detailed judgments about certainty in the evidence for each domain separately and across domains with associated explanations, for example, detailed judgments about the indirectness of the evidence	Judgments about certainty in the evidence as explanations
Certainty, quality, or strength of the evidence	Overall certainty, quality, or strength of the evidence
Number of events and participants in the intervention and control groups	Total number of participants and studies
Importance of outcome	Interpretation and additional comments to facilitate interpretation

## What is new?

### Key findings

- Thus far, there is little guidance for developers of the Grading of Recommendations Assessment, Development and Evaluation (GRADE) evidence tables (Summary of Findings [SoFs] tables and Evidence Profiles [EPs]) about appropriate explanations (also known as footnotes) to optimally justify judgments when reporting their assessment of evidence certainty (also known as quality or confidence) in the context of systematic reviews, health technology assessment (HTA) reports, or guidelines. Explanatory footnotes should be concise, informative, relevant, easy to understand, and accurate.

### What this adds to what was known?

- When using the GRADE approach, transparency of judgments is a key feature to justify interpretation of evidence and to optimally inform decision making. This article presents general and GRADE domain-specific guidance to assist developers of GRADE evidence tables in the creation of optimal explanations, with the ultimate goal of enhancing transparency.

### What are the implications and what should change now?

- Adhering to the general and GRADE domain-specific guidance for developing useful explanations will improve the quality of the available information for users of GRADE evidence tables. This guidance should be incorporated into educational and other material for developers of GRADE evidence tables. SoFs tables should be supported by complete GRADE EPs with more detailed explanations.

results of a randomized trial comparing a new format of SoF tables to the standard format [10]. The second described the results of a systematic survey of explanatory footnotes and comments in SoF tables in Cochrane Reviews and EPs in guidelines from a broad group of organizations [11]. The objectives of this study were to define the essential attributes of useful explanations and to develop specific guidance for providing high-quality explanations for GRADE evidence tables and associated sections in systematic reviews, HTAs and other evidence syntheses. The guidance aims to provide practical suggestions for developing transparent explanations that will make the statements more informative to readers of the tables. Providing this practical guidance responds to strong requests from the large group

of users of the GRADE approach. The work is based on findings of the research reported in the prior two articles [10,11] and indicates that explanations may differ depending on their use in EPs or SoF tables. If a SoF table is chosen as a way to present the key findings, it should be supported by a full EP. These EPs usually provide further details and be placed in an appendix of a systematic review or HTA. Thus, the level of detail of explanations may differ if an SoF table is chosen to present information upfront with a supporting EP.

## 2. Methods

### 2.1. Development of the list of attributes

Based on our work on the previous two articles on SoF tables, we initially identified and selected explanations based on their complexity, type of judgment involved, and appropriateness [10,11]. As part of that work, we rated explanations as “informative, not informative, too detailed, or too open.” Using the inventory of explanatory footnotes and the judgments about them, the lead authors of this article (N.S., A.C.-L., M.L., R.B.-P., R.M., J.B., and H.J.S.) evaluated the merits of each statement by indicating what characteristic of an explanation would have led to a particular rating. This process provided the first list of nine possible attributes that were identified in no specific order: concise, informative, relevant, easy to understand, clear, complete, accurate, nonrepetitive, and interpretable. The lead authors then brainstormed if there were other possible attributes and conducted a brief internet search of literature addressing informative statements and essential characteristics of clear writing. To identify a final list of attributes, the lead authors had multiple rounds of discussion using examples of explanations that demonstrated these desirable attributes. After reviewing 60 explanations published in Cochrane reviews and World Health Organization guidelines, we reached saturation and identified no additional desirable features or attributes that were not addressed by the findings thus far. The redundancy of attributes and overlap led us to agree on a final list of five attributes. For example, the attributes “easy to understand” and “clear” appeared redundant when evaluated with examples because they addressed a similar issue in the explanation. With the reduction to five attributes, we returned to the inventory and identified examples describing these attributes.

### 2.2. Validation of the list of attributes

To ensure completeness of the guidance included in this article, we conducted an iterative revision process with members of the GRADE working group (P.H., T.L., N.O., I.K., D.S., P.G., S.T., D.T., E.A., P.T., and G.G.) who have various editorial roles related to systematic reviews and guidelines. Although we identified a number of examples with important shortcomings and felt that providing all

**Table 1.** Desirable attributes of explanations in Summary of Findings (SoFs) tables and Evidence Profiles (EPs)

Attribute and definition	Key requirements and issues addressed	Examples of acceptable and suboptimal explanations (including examples for differences between SoF tables and EPs)
<p>Concise: The explanation should be brief and free from information that is reported in the main text of the review. As a rule of thumb, an explanation should be not longer than two printed lines.</p>	<ol style="list-style-type: none"> <li>1. Avoid including information that can be included within the table Information about how the outcome was measured or the scale used (in the outcomes column)</li> <li>2. Avoid including information that fits better in the main text Specific details about the analysis or design of included studies. The latter will be covered in the explanation related to risk of bias (if necessary).</li> <li>3. Avoid duplicating information On completion of the table, review all explanations to determine if some explanations could be used multiple times if reworded or combined. Some users may want to keep judgments about the certainty in the evidence separated by domain, in particular in EPs. Because domains may be linked, in specific circumstances, more than one judgment about quality can be included in one explanation.</li> <li>4. Avoid redundant information Do not explain that there are no limitations unless the judgment was challenging and users may wonder why it was not addressed.</li> </ol>	<p>Concise for SoF: Imprecision. The 95% confidence interval includes both no effect and appreciable harm exceeding a minimal important difference.</p> <p>Concise for EP: Very serious imprecision. The 95% confidence interval of the pooled relative effect crosses 1 and includes a 3% absolute reduction in death which might be considered important given the low cost and potential harms of the intervention. The optimal information size to detect a minimal important beneficial effect of the magnitude observed here would require &gt;2,000 patients with 250 events (note: here referring to an outcome associated with fewer events).</p>
<p>Informative: The explanation should provide appropriate information to users regarding (1) judgments; (2) additional sources of information; and (3) interpretation of results</p>	<ol style="list-style-type: none"> <li>1. Judgments When downgrading or upgrading is required, authors should provide the rationale for such decision. In addition, they should provide explanations for borderline decisions or justify judgments when there is some concern about the evidence but the certainty rating is not affected. For details about how to provide appropriate explanations for judgments, see <a href="#">Table 2</a>.</li> <li>2. Additional sources of information This source of information should be related enough to the content of the GRADE table that, in its absence, the autonomy of the table from the main text would be weakened (e.g., where the baseline risk was taken from).</li> <li>3. Knowledge to assist users with interpreting the results This type of explanation should be included when authors anticipate that users would struggle interpreting the results without its inclusion. (e.g., established minimal important difference for a given outcome)</li> </ol>	<p>Informative: Serious indirectness. One trial included children only, one trial adult soldiers, and one trial adults and children (<math>\geq 14</math> yrs). The effect in adults living in the community may be different.</p> <p>Not informative: Risk of bias. Open-label study; not downgraded for this.</p> <p>Informative: Although this was an open-label study, the outcome assessment appeared free from bias because the investigators protected against bias by blinding outcome assessors and reducing the possibility of altering other care influenced by the open-label design.</p> <p>Not informative: The treatment difference was not significant.</p> <p>Informative: The 95% confidence interval includes both no effect and appreciable harm exceeding a minimal important difference.</p> <p>Not informative: <math>RR &gt; 2</math>.</p> <p>Informative: The quality of evidence was upgraded because of a large effect exceeding an RR of 2 with precise estimates from observational studies that did not suffer from risk of bias or other important limitations.</p> <p>Not informative: The control group received placebo.</p> <p>Informative: The control group received placebo in 3 of 5 trials, and the effects were similar in the trials using placebo and</p>

(Continued)

Table 1. Continued

Attribute and definition	Key requirements and issues addressed	Examples of acceptable and suboptimal explanations (including examples for differences between SoF tables and EPs)
Relevant: The explanation should be created keeping in mind the target audience for the GRADE table.	Consider the type of audience your table will inform and how the table will be used for decision making. In EPs, including the cited references for included studies may be of high importance to the guideline panel while meeting to formulate recommendations. However, if the same references are included in an SoF published accompanying a full systematic review, these can be considered as redundant, unnecessary, and irrelevant for users.	those that did not. Therefore, we did not downgrade the quality of evidence. Not informative: Number needed to treat (NNT) = n/a when result is not statistically significant. Relevant: We downgraded two levels because lack of blinding of patients and providers in 4 of 5 studies; it was unclear if allocation was concealed in 2 studies; and only one study clearly used intention to treat analysis. Irrelevant (better placed in methods section of the review): NNT for dichotomous outcomes calculated using NNT calculator ( <a href="http://www.gradepr.org">www.gradepr.org</a> ). Irrelevant: No serious inconsistency. The findings of all the trials are consistent. (Note: this information is irrelevant because it is mandatory to consider this criterion and this does not add additional information when the answer to the question about inconsistency is “no”).
Easy to understand: The explanation should be self-explanatory, not vague, in a way that most users and other target audiences would understand the content provided.	After creating your explanation, read it and determine to what extent it is self-explanatory and clear. Consider having a second author check your explanations. If acronyms are used, make sure the full definition is provided.	Unclear: Confidence intervals include no difference. Easy to understand: Imprecision is present because the width of confidence interval is consistent with both important benefit and harm. Unclear: Although the confidence intervals do not overlap 0.75 and 1.25, the confidence intervals were wide. Easy to understand: Although the confidence intervals do not overlap relative risk estimates of 0.75 and 1.25, the confidence intervals were considered to overlap with the thresholds for decision making. (Note: for EPs, the thresholds should be mentioned). Unclear: According to the results of a trial sequential analysis, there is firm evidence for a beneficial effect of (drug) vs. no placebo or intervention on the (outcome) when the cumulative meta-analysis is adjusted for sparse data and multiple testing on accumulating data. Therefore, there is no risk for random error. Easy to understand for SoF: The confidence intervals are sufficiently narrow for decision making based on statistical analysis and we did not downgrade. Easy to understand for EP: The confidence intervals are sufficiently narrow for decision making and we did not downgrade. This is based a trial sequential analysis in which we adjusted for sparse data and multiple testing.
Accurate: The content of the explanation should be correct and complete.	When the explanation refers to the judgments about the quality of evidence, they should follow the guidance on assessing the	Inaccurate: Heterogeneity is significantly different. Accurate: There was unexplained

(Continued)

Table 1. Continued

Attribute and definition	Key requirements and issues addressed	Examples of acceptable and suboptimal explanations (including examples for differences between SoF tables and EPs)
	certainty in the evidence using the GRADE approach.	<p>inconsistency that was supported by nonoverlapping confidence intervals, high <math>I^2</math> values, and statistically significant heterogeneity of effect estimates.</p> <p>Inaccurate: The number of trials was too few to assess inconsistency.</p> <p>Accurate: The number of studies was small, but no unexplained inconsistency was detected.</p> <p>Inaccurate: Downgraded by 1 for imprecision. The presented data appear highly skewed and could not be pooled.</p> <p>Accurate: Certainty in evidence lowered because of a small number of events leading to wide confidence intervals.</p> <p>Inaccurate: Downgraded (1 level) because the effect was estimated from a single trial. (Note: presence of a single trial is not a reason for imprecision as the trial may be very large, with a sufficient number of events).</p> <p>Inaccurate: Wide confidence intervals indicate significant imprecision of this pooled outcome variable, which causes potential bias. (Note: the imprecision causes random error not bias; the statement should have been avoided or refer to imprecision as a reason for downgrading only).</p>

*Abbreviation:* GRADE, Grading of Recommendations Assessment, Development and Evaluation.

Guidance and examples refer to explanations only and not if the information is useful for other reasons and in other sections in a review, HTA, or guideline.

details of these examples might have made certain aspects clearer, we have removed identifiers.

### 2.3. Development of practical suggestions

We then developed practical suggestions based on the findings of this article, on our experiences supporting editorial groups in Cochrane, systematic reviews, and guideline development, and on feedback from the coauthors who did not develop the attributes we described here (P.H., N.O., T.B., I.K., D.S., P.G., S.T., D.T., E.A., G.G., and P.T.). These practical suggestions are placed in the discussion section as they were not systematically developed but a result of the work on the guidance above.

## 3. Results

### 3.1. Desirable attributes of explanations

We identified that explanations may differ depending on their use in EPs or SoF tables. Although SoF tables are ideally supported by a full EP that provides further details, the latter may be placed in an appendix of a systematic

review or HTA. Thus, the level of detail in explanations may differ if an SoF table is chosen to present information upfront with a supporting EP. Regardless of the format and purpose of presentation, our findings indicate that explanations should have five desirable attributes or characteristics. (Table 1 provides additional guidance to create explanations and examples for each of the attributes.)

#### 3.1.1. Concise

When writing concise explanations, developers should focus on what will be the intended use of the GRADE evidence table. For example, although a GRADE evidence table is presented along with the main text of a systematic review, most evidence tables should be created as if they were a stand-alone source of information. That is, users interested in learning about the effects of a particular intervention would not need to consult the main text to obtain an overview of the findings. Users seeking a deeper understanding of the details related to other aspects of a systematic review or guideline, such as the included studies and review methods, the review search strategy, references, additional tables, or complementary figures as funnel plots, will consult the main text of the review. In other situations, the table is



indeed presented as a stand-alone source of information without an accompanying full text, for example, for EPs used in a guideline or a GRADE evidence to decision framework. In this case, explanations still have to be brief, yet detailed enough to facilitate understanding of the judgments and access to additional information, for example, by including references. Thus, appropriate conciseness depends on if the explanation accompanies an SoF table or an EP; for the latter, greater detail may be required.

In general terms, the conciseness of an explanation relies on three main criteria:

1. Avoid using explanations to present information that is more appropriately presented in the main body of the table. For example, authors of GRADE evidence tables may want to inform users about how an outcome (e.g., quality of life) was measured (e.g., what specific instrument), or the length of follow-up. The outcome column of SoF tables and the header of each outcome for EPs are the appropriate place for providing this information, reducing the amount of text outside the table.
2. Provide information that is not directly needed for making decisions in the main text of the review or guideline and not in the explanations. For example, a specific description of statistical methods is better included in the methods section of the review than in the table.
3. Avoid duplicating information. Sometimes, one explanation can be used to justify more than one judgment in the table, for more than one outcome. Authors may be inclined to keep judgments on the certainty in the evidence separated by domain; we believe this is not optimal when one judgment addresses related domains, for example, a judgment about imprecision and inconsistency.
4. Avoid redundant information. Do not explain that there are no limitations unless the judgment was challenging and users may wonder why it was not addressed, for example, in EPs.

### 3.1.2. Informative

An explanation is informative when it provides appropriate data to decision makers and other stakeholders. There are three types of informative explanations (Table 1). First, explanations can inform about the rationale for judgments about the certainty in the evidence. This includes decisions to not upgrade or downgrade the evidence when raters face borderline decisions or justify not rating down despite concern about the evidence. Second, explanations can be used to link the evidence to additional sources of information. Third, informative explanations can assist users with interpreting the results presented in the table.

### 3.1.3. Relevant (to the identified target audience)

Explanations should be developed with the target audience for the GRADE evidence table in mind. EPs should

provide comprehensive explanations about judgments related to rating certainty in the evidence or issues related to the baseline risk for audiences who are expecting these additional details. Consider a situation in which an EP is part of the meeting documents that inform the process of formulating recommendations by a panel. The expectations about the content of explanations are different compared to a situation in which an SoF table summarizes the results of a systematic review for other users, such as a clinician reading the review (see Table 2 for examples). Panel members often like to see the source of information for the outcomes in an EP to evaluate if the relevant evidence is included. If the original EP is used for a guideline focusing on a different population than the systematic review, this should be considered in the judgment about indirectness. This judgment may differ from the original judgment of the systematic review. On the other hand, for an SoF table based on an EP in a systematic review, reporting the source of information is redundant.

### 3.1.4. Easy to understand

An effective explanation should be self-explanatory for the target audience. This is a key characteristic given the broad range of users who access SoF tables and EPs published in systematic reviews and guidelines, in particular for SoF tables that may be used by lay audiences. Explanations that are easy to understand are not vague, use simple terms to describe a given judgment, and do not require users to go to the main text of the review or guideline for basic information.

### 3.1.5. Accurate

Although it is obvious that statements must be accurate, an explanation can have all the attributes listed previously and still be inaccurate because its content does not follow the GRADE guidance or the information provided is incorrect. For example, the statement “Inconsistency was not assessed because only one study was identified” may be concise, considered informative, relevant to the context, and easy to understand, but it is wrong according to GRADE guidance. When only one study is included in a review of the evidence, the rating for inconsistency becomes “not serious” indicating there is no inconsistency (note: this is frequently misunderstood by raters) [14]. Other examples include that judgments about imprecision should not be based on statistical significance alone because they depend on other factors such as the number of events and if the relative or absolute measure of effect is considered [15]. Inaccurate explanations cast doubt on the information presented in the GRADE evidence table, affecting its credibility. We advise developers not only to follow the guidance presented here but also to consult the series of articles in the Journal of Clinical Epidemiology and the GRADE handbook in the GRADEpro software ([www.grade.pro](http://www.grade.pro)) to assess certainty in the evidence and move from evidence to decisions [5,16,17].

**Table 2.** Content of explanation for reporting judgments on the certainty in the evidence assessment

GRADE domain	Relevant content	Example
Criteria for downgrading		
Risk of bias (also known as limitations in detailed study design and execution)	<ul style="list-style-type: none"> <li>- Whether there are serious or very serious issues of risk of bias</li> <li>- Proportion/number of studies showing shortcomings with any of the risk of bias issues by choosing the appropriate risk of bias tool depending on if randomized or nonrandomized studies were used</li> <li>- Contribution of the studies to the pooled estimates</li> </ul>	<p>Risk of bias: Studies that carried large weight for the overall effect estimate rated as high risk of bias due to lack of concealment of randomization and lack of blinding (for SoF only).</p> <p>Studies that carried large weight for the overall effect estimate rated as high risk of bias due to lack of [concealment] in [3] out of [5] studies and [lack of blinding] in [2] out of [5] studies. (For EP: information in [] can be substituted based on the example.)</p>
Imprecision	<ul style="list-style-type: none"> <li>- Whether there is serious or very serious issues of imprecision</li> <li>- Interpretation of the limits of the confidence interval</li> <li>- Determine whether the optimal information size is met</li> </ul>	Serious (or very serious) imprecision. 95% CI is consistent with the possibility for important benefit and large harm exceeding a minimal important difference, including only [] events in total.
Inconsistency	<ul style="list-style-type: none"> <li>- Whether there are serious or very serious issues of inconsistency</li> <li>- Unexplained heterogeneity identified by analysis of point estimates and confidence interval overlap, statistical test (chi-square), and statistical estimates (<math>I^2</math>)</li> </ul>	Serious (or very serious) inconsistency. Unexplained inconsistency, with point estimates widely different and confidence intervals not overlapping ( $P$ -value chi square = []; $I^2$ = [] %)
Indirectness	<ul style="list-style-type: none"> <li>- Whether there are serious or very serious issues of indirectness</li> <li>- Any substantial difference between the identified evidence and the original question of the review regarding: patients, interventions, comparisons, outcomes to the extent that these differences would question the obtained effect estimate</li> <li>- Authors should complete the indirectness table assesses in detail indirectness related to the review or guideline question PICO: population, intervention, comparison, and outcome [12,13] (<a href="http://www.gradepr.org">www.gradepr.org</a>)</li> </ul>	Serious (or very serious) indirectness. Patients included in the studies have [different condition], and differ importantly from the [question]. Studies also used different [doses of drug] compared to the [question].
Publication bias	<ul style="list-style-type: none"> <li>- Whether publication bias is undetected or suspected</li> <li>- Interpretation of funnel plot</li> <li>- Comprehensiveness of the search strategies and methods to identify all available evidence</li> <li>- Presence of small (often positive) studies with for profit interest</li> </ul>	Publication bias was suspected because the included studies were small and the funnel plot shows asymmetry.
Criteria for upgrading		
Large effect	<ul style="list-style-type: none"> <li>- Whether a large effect or association is present and the time frame of exposure necessary to achieve the effect</li> <li>- Explicit description of the magnitude of effect considered as large</li> </ul>	Large effect based on well-done observational studies without important risk of bias or other limitations showing an OR = _____ (95% CI: _____, _____).
Dose–response gradient	<ul style="list-style-type: none"> <li>- Whether the studies provide evidence of a dose–response gradient between intervention or exposure and outcome</li> <li>- Explicit description of the intervention’s or exposure’s thresholds related to changes (improvement or reduction) in the outcome</li> </ul>	Clear dose–response gradient. RR with the intervention _____ (95% CI: _____, _____) with doses less than _____ and RR of _____ (95% CI: _____, _____) with doses larger than _____.
Opposing plausible residual confounding and bias	<ul style="list-style-type: none"> <li>- Whether the studies provide evidence of all plausible confounders or biases against the detected effect or association, when the later is still detected</li> <li>- Whether the studies provide evidence of all plausible confounders or biases in favor of detecting an effect or association, when the later is not detected</li> <li>- Explicit description of the mechanism for which confounders or biases may be reducing or increasing the observed effect or association</li> </ul>	<p>Confounding and/or biases against the detected effect or association (description of the mechanism should be provided).</p> <p>Confounding and/or biases in favor of detecting an effect or association not found (description of the mechanism should be provided).</p>



**Table 3.** Guidance for providing useful explanations in Summary of Findings (SoFs) Tables and Evidence Profiles (EPs)

1. Much of the information you want to communicate to readers can be entered directly into the table and an explanation may not be necessary (e.g., information about the duration of follow-up or the scale used).
2. Generally, do not cite references in the explanations section, unless the GRADE evidence table will be used as a stand-alone EP, for example, in a guideline panel meeting or for providing information about baseline risks.
3. The source of information about the baseline risks used to calculate absolute effects should be provided.
4. Depending on the type of table (SoF, EP, stand alone vs. part of a review) explanations may be more or less concise as readers can refer to the review for details.
5. On completion of the table, review all of the explanations to determine if some could be referred to multiple times if reworded or combined.
6. Provide reasons for upgrading and downgrading the evidence (see domain-specific guidance in the following) and use GRADEpro software to adhere to GRADE guidance.
7. You should consider describing whether you determined that the body of evidence for a particular outcome has serious or very serious issues for the affected domain. Thus, it may be useful to indicate the number of levels for downgrading (e.g., downgraded by one level for risk of bias), but you should avoid repetition of what is in the table and the impression of formulaic or algorithmic reporting. In EPs, this information is already in the cells of the table.
8. Although explanations about the certainty in the evidence are primarily required when they alter the certainty, you could add an explanation to explain when you have not altered the certainty in the evidence, but this decision may be questioned by others. This will help with understanding reasons for disagreement.
9. Remember the table is a summary of the findings not a description of the methods of the review (e.g., do not describe the reasons for the statistical analysis).
10. Results for outcomes that could not be combined statistically in a meta-analysis (i.e., narrative outcomes) can be entered directly into the GRADE evidence table under the results column. An explanation may not be necessary to communicate those results. If you consider that your audience would benefit from adding complementary estimates of treatment effect (e.g., number needed to treat for benefit and harm, risk difference expressed as percentage, continuous outcome expressed in minimal important difference units), these can be included in the comment column.
11. Use the information presented in the explanations in SoF about the GRADE process to develop other key parts of the systematic review, including summary versions and discussion section [8].

#### Domain-specific guidance for writing useful explanations

##### Risk of bias

1. Qualify the number of studies, or the amount of information that they provide in the meta-analysis, which were at high risk of bias and indicate for which criterion.
  - a. Use terms such as majority, minority, all, some, or none; or indicate the number as X/X studies.
  - b. For randomized controlled trials, refer to the specific criteria including allocation concealment, selective outcome reporting, and so forth. For nonrandomized studies, refer to the criterion in the tool used (e.g., using the ACROBAT-NRS tool). In some instances, also indicate the contribution of the studies to the estimates.
  - c. Indicate if the effect of the risk of bias was tested in a sensitivity analysis.
2. Information about study design may be included in the explanations, in particular, in SoF when different study designs are included. However, this information is included in the EP.

##### Imprecision

1. Indicate whether the sample size or number of events does not meet the optimal information size as calculated, or the “rules of thumb” (e.g., >400 events). Avoid reference to the number of studies as a reason for imprecision.
2. Indicate whether the confidence intervals include the possibility of a small or no effect AND important benefit or harm. If the threshold of important benefit is known, provide the numerical value.
3. Avoid reporting the result as statistically or nonstatistically significant.

##### Inconsistency

1. Indicate if the judgment of inconsistency is based on statistical tests ( $I^2$ ,  $\chi^2$ , tau), overlap of confidence intervals, or similarity of point estimates.
2. If inconsistency is based on  $I^2$ , consider describing it as considerable, substantial, moderate, or not important.
3. If applicable, describe whether heterogeneity was explored in subgroup analyses by PICO (patients, intervention, comparison, outcome), and indicate other potential reasons for the heterogeneity.
4. In the case of a single study for an outcome, do not indicate that inconsistency is “not applicable.” Indicate that there is none.

##### Indirectness

1. Describe if indirectness is due to the elements of PICO.

##### Publication bias

1. Indicate the reason publication bias is detected (e.g., asymmetrical funnel plot, small studies with positive results, suspected selective availability of data from published, or unpublished studies).

##### Upgrading

1. In SoF compared to EP, the reasons for upgrading are not provided, therefore write the specific reason: due to large effect; a dose–response gradient; or plausible confounding increases confidence.
2. In both SoF and EP, indicate the specific reason. For large effects, report if relative effect is >2 or >5. For dose–response gradients, describe the level of intervention and effect on outcome. For confounding, describe the effect of the confounding factor on the estimate.

---

*Abbreviation:* GRADE, Grading of Recommendations Assessment, Development and Evaluation.

### 3.2. What type of information should be included in the comment column?

Research conducted to improve understanding of GRADE evidence tables found the best use of the comment column was uncertain. Some users preferred less crowded tables suggesting that the comment column was superfluous and might best be omitted. Other users mentioned that the statements in the comment column may be helpful to interpret information and require a prominent place [10].

In line with these results, we suggest that the comment column is used when developers need to help readers interpret review findings [9,18]. Unlike explanations, the comments column presents information that aids with interpretation but is not essential for understanding. Developers could present alternative effect measures such as the number needed to treat for benefit or harm for each outcome, a continuous outcome in minimal important difference units, or a hazard ratio [18]. Finally, help with interpretation in the comment column may be provided by indicating if an effect is large, small, or absent and a narrative about the certainty in this effect.

### 3.3. Standardized explanations to advise users of SoF tables and EPs

Although some explanations are unique and specific to a particular situation, there are common themes and our intent is to provide guidance for common issues. And, although it is not our intention to restrict the use of explanations only to the examples described here, we strongly encourage developers of GRADE evidence tables to make use of standard explanations. Table 2 provides examples of the minimum content and standard explanation.

### 3.4. Where and how can information presented in explanations be used in systematic reviews?

Information used to justify downgrading or upgrading decisions can help researchers improve the consistency of interpretation and reporting across different parts of a systematic review. Reporting guidance for systematic reviews increasingly emphasizes the need to present findings both in terms of the results and the certainty in effects. Both need to be reported across full text and summary versions of systematic reviews [19–22]. Inconsistency can arise simply from not reusing information that is provided in SoF tables in the main body of a systematic review. Undertaking a GRADE assessment should not be regarded solely in terms of the process of preparing an SoF table [8,23] but as an aid to interpret the key results of the systematic review. This interpretation will inform the discussion, main conclusions, and the summary versions or products of the review. Given the need for brevity and plain language in summary versions or products of a review, it may not always be appropriate or useful to provide a detailed breakdown of downgrading decisions. However, when

summarizing the key findings of the review, a comment on how the certainty in the evidence was derived should be included and that can be based on the explanations.

## 4. Discussion

Explanations are a critical component of GRADE evidence tables, but their use varies and could be improved. Based on our user testing that informed the RCT described in the first article in this series and the critical appraisal of explanations published in Cochrane systematic reviews and clinical practice guidelines [10,11], we now provide detailed guidance for these explanations. Our analysis allowed us to describe the desirable features of explanations, areas in which guidance is needed, and minimum content for specific explanations, which we provide in Tables 2 and 3.

The strength of this work includes a comprehensive and systematic analysis of a wide range of SoF tables and EPs published in Cochrane systematic reviews and clinical practice guidelines [11]. These GRADE tables included both quantitative and narrative syntheses of results, which suggests that our guidance for standardized explanations can be applied to both quantitatively and qualitatively synthesized data. To inform developers of GRADE evidence tables, we also provide a description of the desirable attributes for explanations. The involvement of a group of authors and editors with experience in developing and reviewing hundreds of SoF tables and EPs provides another strength.

The work has some limitations. First, our standardized text (Tables 1 and 2) cannot cover all eventualities and some variability will remain. We hope that the guidance given in Table 1 will reduce this variability without imposing rigidity. Second, although the desirable attributes listed in Table 1 seem to have face validity and would, we think, improve the explanations given in GRADE evidence tables, large-scale confirmation of this by users and developers is still required. Third, the GRADE evidence tables in our database did not include tables that are currently provided interactively (i.e., interactive SoFs). Some may argue that explanations could be less concise if provided electronically, but we believe all of the attributes, including conciseness, pertain to interactive GRADE tables and explanations. Fourth, we have not provided examples for explanations related to GRADE tables for prognostic or test accuracy questions. The described attributes apply to those tables, but more work is needed to provide examples and detailed guidance in those contexts.

When developers of GRADE evidence tables are using explanations to optimize transparency about judgments, they should keep in mind that appropriate explanations are concise, informative, relevant, easy to understand, and accurate in their content. Although Table 1 provides advice to transfer these desirable characteristics into practice, we

believe the practical suggestions in Table 3 respond to requests from those applying the GRADE approach in GRADE workshops and other settings.

Explanations in GRADE evidence tables provide information that can also be used to summarize (limitations of) the evidence in other sections of the systematic review. For example, researchers preparing systematic reviews may find it useful to draw on downgrading decisions presented in explanations to justify their assertions about the certainty in the evidence in the discussion section of a review. They could also use these decisions to identify implications for research, for example, if evidence was downgraded because of attrition bias, how should future studies address this? Guidance for implications for research is also provided in the Cochrane Handbook of Systematic Reviews [8]. Editors handling articles of systematic reviews may find it useful to assess the consistency of reporting by comparing assertions about the quality of evidence or conclusions made in the text with decisions provided in explanations of GRADE evidence tables.

## 5. Conclusions

Although SoF tables and EPs are now widely used, guidance on how to provide explanations as one of the key features of the tables has been lacking. We have analyzed explanations in Cochrane reviews and practice guidelines that included SoF tables or EPs. Although SoF tables are abbreviated versions of EPs with shorter explanations and intended for different target audiences, their backbone is a full EP which should be provided to inform an SoF table. Our work suggests attributes for explanations that enhance the development and interpretation of these tables and will aid with relevant sections in the main body of an evidence synthesis such as a systematic review or HTA.

## Acknowledgments

The GRADE Center at McMaster University and the Methods Innovation Fund of the Cochrane Collaboration funded this study. Neither of these institutions played a role in the planning, conduct, or publishing the study findings except through the authors' affiliation.

## References

- [1] Schünemann HJ, Best D, Vist G, Oxman AD, GRADE Working Group. Letters, numbers, symbols and words: how to communicate grades of evidence and recommendations. *CMAJ* 2003;169: 677–80.
- [2] 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.
- [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] Mustafa RA, Santesso N, Brozek J, Akl EA, Walter SD, Norman G, et al. The GRADE approach is reproducible in assessing the quality of evidence of quantitative evidence syntheses. *J Clin Epidemiol* 2013;66:736–42. quiz 42 e1–5.
- [5] 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.
- [6] 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.
- [7] Guyatt GH, Thorlund K, Oxman AD, Walter SD, Patrick D, Furukawa TA, et al. GRADE guidelines: 13. Preparing summary of findings tables and evidence profiles-continuous outcomes. *J Clin Epidemiol* 2013;66:173–83.
- [8] Schünemann HJ, Oxman AD, Vist GE, Higgins JPT, Deeks JJ, Glasziou P, et al. Chapter 12: interpreting results and drawing conclusions. In: Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.10. The Cochrane Collaboration; 2011. (updated March 2011). Available at <http://www.cochrane-handbook.org>. Accessed November 25, 2015.
- [9] Schünemann HJ, Oxman AD, Higgins JPT, Vist GE, Glasziou P, Guyatt GH. Chapter 11: presenting results and ‘Summary of findings’ tables. In: Higgins JPT, Green S, editors. *Cochrane Handbook for Systematic Reviews of Interventions* Version 5.10. The Cochrane Collaboration; 2011. (updated March 2011). Available at <http://www.cochrane-handbook.org>. Accessed November 25, 2015.
- [10] Carrasco-Labra A, Brignardello-Petersen R, Santesso N, Neumann I, Mustafa R, Mbuagbaw L, et al. Improving GRADE evidence tables: a randomized trial shows improved understanding of content in Summary-of-Findings tables with a new format. *J Clin Epidemiol* 2016;74:7–18.
- [11] Langendam M, Carrasco-Labra A, Santesso N, Brignardello-Petersen R, Ventresca M, Heus P, et al. Improving GRADE evidence tables part 2: a systematic survey of explanatory notes shows more guidance is needed. *J Clin Epidemiol* 2016;74:19–27.
- [12] Schunemann HJ. Methodological idiosyncracies, frameworks and challenges of non-pharmaceutical and non-technical treatment interventions. *Z Evid Fortbild Qual Gesundheitswes* 2013;107(3):214–20.
- [13] Schünemann HJ, Tugwell P, Reeves BC, Akl EA, Santesso N, Spencer FA, et al. Non-randomized studies as a source of complementary, sequential or replacement evidence for randomized controlled trials in systematic reviews on the effects of interventions. *Res Synth Methods* 2013;4(1):49–62.
- [14] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. *J Clin Epidemiol* 2011;64:1294–302.
- [15] 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.
- [16] Balslem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol* 2011;64:401–6.
- [17] Andrews JC, Schunemann HJ, Oxman AD, Pottie K, Meerpohl JJ, Coello PA, et al. GRADE guidelines: 15. Going from evidence to recommendation—determinants of a recommendation’s direction and strength. *J Clin Epidemiol* 2013;66:726–35.
- [18] Akl EA, Schunemann HJ. Routine heparin for patients with cancer? One answer, more questions. *New Engl J Med* 2012;366:661–2.
- [19] McIlwain C, Santesso N, Simi S, Napoli M, Lasserson T, Welsh E, et al. Standards for the reporting of Plain Language Summaries in new Cochrane Intervention Reviews (PLEACS). *Methodological Expectations of Cochrane Intervention Reviews (MECIR)* [Internet]. February, 2003 May, 2015. Available at [http://editorial-unit.cochrane.org/sites/editorial-unit.cochrane.org/files/uploads/PLEACS\\_0.pdf](http://editorial-unit.cochrane.org/sites/editorial-unit.cochrane.org/files/uploads/PLEACS_0.pdf).

- [20] Chandler J, Churchill R, Higgins J, Lasserson T, Tovey D. Methodological standards for the conduct of new Cochrane Intervention Reviews. Methodological Expectations of Cochrane Intervention Reviews (MECIR) [Internet]. December, 2013 May, 2015. Available at [http://editorial-unit.cochrane.org/sites/editorial-unit.cochrane.org/files/uploads/MECIR\\_conduct\\_standards2.3\\_02122013.pdf](http://editorial-unit.cochrane.org/sites/editorial-unit.cochrane.org/files/uploads/MECIR_conduct_standards2.3_02122013.pdf).
- [21] Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *J Clin Epidemiol* 2009;62:1006–12.
- [22] Moher D, Shamseer L, Clarke M, Ghersi D, Liberati A, Petticrew M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst Rev* 2015;4:1.
- [23] Lasserson T, Santesso N, Cumpston M, Marshall R, Ní Ógáin O. Incorporating GRADE in Cochrane Reviews: feedback from the CEU screening programme 2015. Available at: [http://training.cochrane.org/sites/training.cochrane.org/files/public/uploads/resources/downloadable\\_resources/English/Incorporating%20GRADE%20in%20Cochrane%20Reviews%20PDF.pdf](http://training.cochrane.org/sites/training.cochrane.org/files/public/uploads/resources/downloadable_resources/English/Incorporating%20GRADE%20in%20Cochrane%20Reviews%20PDF.pdf). Accessed November 15, 2015.