Journal of Clinical Epidemiology 93 (2018) 36–44

Journal of Clinical Epidemiology

Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis

Romina Brignardello-Petersen[a,b], Ashley Bonner[a], Paul E. Alexander[a,c], Reed A. Siemieniuk[a,d], Toshi A. Furukawa[e,f], Bram Rochwerg[a,g], Glen S. Hazlewood[h,i], Waleed Alhazzani[a,g], Reem A. Mustafa[a,j], M. Hassan Murad[k], Milo A. Puhan[l,m], Holger J. Schünemann[a], Gordon H. Guyatt[a,*], For the GRADE Working Group

[a]*Department of Health Research Methods, Evidence and Impact, McMaster University, 1280 Main St W, Hamilton, Ontario L8S 4K1, Canada*
[b]*Evidence Based Dentistry Unit, Faculty of Dentistry, Universidad de Chile, Sergio Livingstone Pohlhammer 943, Independencia, Santiago, Chile*
[c]*The Infectious Diseases Society of America, 1300 Wilson Boulevard, Suite 300, Arlington, VA 22209, USA*
[d]*Department of Medicine, University of Toronto, 190 Elizabeth Street. R. Fraser Elliott Building, 3-805, Toronto, Ontario M5G 2C4, Canada*
[e]*Department of Health Promotion and Human Behavior, Kyoto University Graduate School of Medicine/School of Public Health, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan*
[f]*Department of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501, Japan*
[g]*Department of Medicine, McMaster University, 1280 Main St W, Hamilton, Ontario L8S 4L8, Canada*
[h]*Department of Medicine, Cumming School of Medicine, University of Calgary, 3330 Hospital Drive NW, Calgary, Alberta T2N 4N1, Canada*
[i]*Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, 3330 Hospital Drive NW, Calgary, Alberta T2N 4N1, Canada*
[j]*Division of Nephrology and Hypertension, Department of Medicine, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, KS 66160, USA*
[k]*Mayo Clinic Evidence Based Practice Center, Harwick Building, Room 2-54, Rochester, MN 55905, USA*
[l]*Epidemiology, Biostatistics und Prevention Institute (EBPI), University of Zurich, Hirschengraben 84, Zurich 8001, Switzerland*
[m]*Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 North Wolfe Street, W6508, Baltimore, MD 21205, USA*

Accepted 1 October 2017; Published online 17 October 2017

Abstract

This article describes conceptual advances of the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) working group guidance to evaluate the certainty of evidence (confidence in evidence, quality of evidence) from network meta-analysis (NMA). Application of the original GRADE guidance, published in 2014, in a number of NMAs has resulted in advances that strengthen its conceptual basis and make the process more efficient. This guidance will be useful for systematic review authors who aim to assess the certainty of all pairwise comparisons from an NMA and who are familiar with the basic concepts of NMA and the traditional GRADE approach for pairwise meta-analysis. Two principles of the original GRADE NMA guidance are that we need to rate the certainty of the evidence for each pairwise comparison within a network separately and that in doing so we need to consider both the direct and indirect evidence. We present, discuss, and illustrate four conceptual advances: (1) consideration of imprecision is not necessary when rating the direct and indirect estimates to inform the rating of NMA estimates, (2) there is no need to rate the indirect evidence when the certainty of the direct evidence is high and the contribution of the direct evidence to the network estimate is at least as great as that of the indirect evidence, (3) we should not trust a statistical test of global incoherence of the network to assess incoherence at the pairwise comparison level, and (4) in the presence of incoherence between direct and indirect evidence, the certainty of the evidence of each estimate can help decide which estimate to believe.

Keywords: GRADE; Quality of evidence; Network meta-analysis; Indirect comparisons; Certainty of evidence; Confidence in estimates of effect

1. Introduction

In 2014, the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) working group presented guidance to evaluate the certainty of the evidence

Conflict of interest: None.

* Corresponding author. 1280 Main St West, Hamilton, Ontario L8S 4L8, Canada. Tel.: 905.521.9140x22900; fax: 905.524.3841.
E-mail address: guyatt@mcmaster.ca (G.H. Guyatt).

https://doi.org/10.1016/j.jclinepi.2017.10.005

(confidence in evidence, quality of evidence) from network meta-analysis (NMA). [1] This guidance represented a response to the need for establishing the certainty of the evidence for each paired comparison within an NMA and the desirability of implementing widely used GRADE criteria [2] to inform those judgments [3].

The application of GRADE's approach to rate the certainty of the evidence in an NMA may appear onerous in networks with many interventions. While in the simplest network with only three treatments (e.g., treatments A, B, and C), researchers must undertake the certainty assessment three times (i.e., they must address the direct, indirect, and network estimates for A vs. B, A vs. C, and B vs. C), the requirement in a network with 6 treatments is 15 assessments and in a network with 12 treatments 66 assessments. Moreover, the assessment requires repetition for each outcome of interest.

The application of the GRADE approach to a number of NMAs [4–8] in the 3 years since the original guidance publication has led to advances that have strengthened the conceptual basis, dealt with challenges that have arisen in applying the approach, and—most relevant to the volume of work required in applying the GRADE approach—may make the process of assessing the certainty of the evidence more efficient. In this article, we describe these advances and their rationale and provide illustrative examples. We focus on guidance for systematic reviewers who aim to rate the certainty of the evidence of all the pairwise comparisons from an NMA, regardless of whether there is high certainty from traditional direct comparisons to inform clinical decision-making. The discussion assumes a familiarity with the basic concepts of indirect evidence and NMA and with the GRADE approach to rating certainty of evidence for bodies of evidence in conventional paired comparison meta-analysis and is restricted to NMAs of randomized trials. This article describes official guidance from the GRADE working group.

2. Assessing the certainty of the evidence from NMA—the GRADE approach

To assess the certainty associated with evidence from NMA, we must consider all the contributing evidence. This includes evidence from trials directly comparing any two interventions of interest—direct evidence—and the evidence from trials that inform an indirect comparison of the two interventions through one or more common comparators—indirect evidence. According to the GRADE approach, the certainty of the evidence from a conventional meta-analysis of randomized trials comparing two alternatives for a specific outcome is based on considerations of risk of bias (limitations in the study design and execution), inconsistency, imprecision, indirectness, and publication bias. [2] Limitations in any of these domains result in rating down the certainty of the evidence from high to moderate, low, or very low certainty.

Within an NMA, the certainty of evidence will almost invariably differ across the paired comparisons being considered. For this reason, the assessment of the certainty of the evidence must be done at the paired comparison level.

In brief, the previously published GRADE approach for assessing the evidence for a specific outcome and comparison in an NMA requires following four steps: [1] (1) presenting the direct and indirect estimates of effect for the pairwise comparison, (2) rating the certainty of both of these estimates, (3) presenting the network estimate for the pairwise comparison, and (4) rating the certainty of the network estimate, based on the ratings of the direct and indirect estimates and the assessment of coherence (i.e., extent of similarity of direct and indirect estimates). For rating the certainty of the indirect estimates, reviewers should focus their assessment on the most-dominant first-order loop. The original GRADE guidance to assess the certainty in estimates from NMA presents details and rationale for each of the steps. [1].

Reviewers and clinicians may wonder why one should bother with rating a network estimate when the certainty of the direct evidence is high. When we undertake an NMA, we proceed under the assumption that it is desirable to use the NMA estimate unless there are compelling reasons to not do so. Further, even if high certainty direct evidence is available, indirect evidence may nevertheless enhance the certainty of the network estimate, if it is coherent with the direct evidence, by further narrowing the confidence interval or by enhancing the applicability (i.e., directness) of the body of evidence. This is because the certainty of evidence is a continuum, and dividing it into four categories of high, moderate, low, and very low is simply a matter of convenience. Further, if direct and indirect estimates are coherent, one can combine lower certainty indirect evidence with higher certainty direct evidence without compromising overall certainty.

Think, for instance, of a visual analogue scale from 0 to 100, in which higher numbers represent higher certainty. Reviewers can find themselves rating estimates as having high certainty when, were they to use the continuous scale,

Table 1. Advances of the GRADE approach for rating the certainty of estimates of effect from NMA

Strategies to achieve efficiency
Consideration of imprecision is not necessary when rating the direct and indirect estimates to inform the rating of NMA estimates.
There is no need to rate the indirect evidence when the certainty of the direct evidence is high and the contribution (i.e., the relative weight in the network estimate) of the direct evidence to the network estimate is at least as great as that of the indirect evidence.
Other relevant considerations
We should not trust a statistical test of global incoherence of the network to assess incoherence at the pairwise comparison level.
In presence of incoherence between direct and indirect evidence, the certainty of the evidence of each the estimates can help decide which estimate to believe.

Abbreviations: GRADE, Grading of Recommendations Assessment, Development, and Evaluation; NMA, network meta-analysis.

their visual analogue scale rating would be 80. Although this is high certainty, it would still be desirable to achieve even greater certainty. If adding the indirect evidence in the NMA narrows the confidence interval and thus moves to even higher certainty (for instance, a VAS rating of certainty of 90), it is desirable to do so.

3. Conceptual advances in the assessment of the certainty of the evidence from NMA

Herein, we describe two modifications of GRADE guidance that may enhance efficiency of the GRADE process. One relates to assessment of imprecision and the other to the possibility of omitting rating of the certainty of the indirect evidence (Table 1).

3.1. Consideration of imprecision is not necessary when rating the direct and indirect estimates to inform the rating of the network estimates

In the GRADE approach, imprecision is one of five domains considered in rating the certainty of the evidence [2]. In the context of clinical practice guidelines, assessing imprecision requires making a judgment of whether the upper and lower boundaries of the confidence interval (we will use this term throughout the article to refer to the confidence intervals obtained using frequentist approaches and credible intervals obtained using Bayesian approaches to NMA) of an estimate of effect would lead to the same clinical action [9]. If the clinical action would not differ, one need not rate down for imprecision; if it would, rating down for imprecision is required.

Consider an NMA in which, for a particular paired comparison, direct and indirect estimates contribute equally to the NMA estimate, and the certainty rating from both estimates is moderate. Following GRADE guidance, the rating of the network estimate should also be moderate certainty. Indeed, if one is rating down both the direct and indirect estimates for any of risk of bias, inconsistency, indirectness, or publication bias, this will be the case.

Consider, however, that the rating down of either direct or indirect estimates has been because of imprecision (i.e., there are no serious problems with any of the other GRADE domains). Because both direct and indirect evidence contribute substantially to the network estimate, the confidence interval of the network estimate may be appreciably narrower than that of either the direct or indirect estimate. If this is the case, the rating of the network estimate may not be moderate, but rather, high (no problems in other domains and no problems in imprecision).

Invoking this logic need not require that precision be the only problematic domain. Consider a situation in which the higher certainty of the direct or indirect evidence is rated down for both imprecision and one of the other four domains. The reviewer would then rate the certainty of the network estimate as low. But if combining the direct and indirect narrows the confidence interval sufficiently that there is no longer any need to rate down for imprecision, then the certainty in the network estimate would move from low to moderate.

This implies that for every network, after rating the direct and indirect certainty and choosing the higher of the two for the rating of network certainty, the reviewer must evaluate the precision of the network estimate. If the network estimate is sufficiently precise (i.e., one would not rate down for imprecision on the basis of the network estimate), the reviewer must go back and check if imprecision is one of the issues responsible for the rating down of the higher certainty of the direct and indirect evidence. If the answer is yes, the reviewer must now rate up one from the higher of the direct and indirect evidence.

There is a way around this somewhat circuitous logic. Even if they are informing the rating of the network estimate, the other four GRADE domains cannot be assessed directly from review of the network estimate. Thus, assessment of risk of bias, inconsistency, indirectness, and publication bias requires separate consideration of the direct and indirect evidence. For the indirect evidence, these four domains are evaluated separately in each of the direct comparisons that contribute to the indirect estimate. In contrast, one can obtain the precision of the network estimate directly, simply by considering the confidence interval around the network estimate. The solution to this issue is therefore to not bother with the rating of the precision of the direct and indirect evidence. The reviewer can rate the direct and indirect evidence on the basis of the other four domains, while reserving the rating of precision to the certainty of the network estimate.

To illustrate, we will use as an example the NMA that compared agents for preventing stress ulcers in mechanically ventilated critically ill patients (Manuscript submitted for publication). The authors considered four treatments: histamine-2 receptors antagonists (H2RAs), proton pump inhibitors, sucralfate, and placebo. In the H2RA vs. sucralfate comparison, for the outcome of pneumonia, the direct evidence showed an odds ratio (OR) of 1.32, with a 95% confidence interval (CI): 0.98, 1.77, mandating rating down for imprecision (if the OR of 1.77 was true, one would hesitate to prescribe H2RA on the basis of pneumonia risk; if the OR was 0.98, this would not be the case) (Fig. 1). The reviewers rated down this direct evidence for risk of bias, but not for inconsistency, indirectness, or publication bias; therefore, they rated the direct evidence as low certainty due to risk of bias and imprecision. The OR for the network estimate showed a relative effect of OR = 1.30, with a 95% CI: 1.08, 1.58. With this narrower confidence interval, considering only the issue of imprecision, one would now confidently infer that H2RAs increase the incidence of pneumonia. Thus, the reviewer might reasonably conclude that there is no need to rate down the network estimate for imprecision, resulting in an overall rating of moderate certainty evidence.

We conclude, therefore, that the assessment of imprecision of the direct and indirect estimates is typically not necessary to inform the network estimate certainty. Rather, reviewers can base the imprecision rating on the confidence

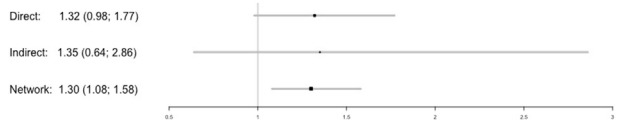

Fig. 1. Direct, indirect, and network estimates of H2RA vs. sucralfate.

interval around the network estimate itself. Systematic review authors who want to gain a better understanding of the relationship between direct and indirect estimates can still assess imprecision when making the separate certainty ratings. Those who are keen on maximum efficiency can, however, refrain from so doing.

3.2. There is no need to rate the indirect evidence when the certainty of the direct evidence is high, and the contribution of the direct evidence to the network estimate is at least as great as that of the indirect evidence

It was implicit in the original GRADE guidance for NMA that investigators undertake the assessment of the certainty of the direct and indirect estimates to inform the certainty of the network estimate. The final guidance did not, however, consider the possibility that, to achieve optimal efficiency of the GRADE rating process, one might sometimes be able to forego rating the indirect evidence. Acknowledging that, in the context of NMA, the direct and indirect evidence ratings are not important in themselves, but only to support the rating of the network estimate, allows the possibility of omitting rating the indirect evidence. When possible, this makes the process more efficient.

GRADE guidance specifies that the rating of the network estimate is based on the higher of the direct and indirect evidence ratings. For example, if the direct estimate has high certainty and the indirect estimate has moderate certainty, the network estimate rating will be high certainty. The reason is that, if the direct and indirect estimates are coherent (i.e., they do not substantially differ), the estimate with the lower certainty would not, relative to the estimate with the higher certainty, introduce bias.

Another way of expressing this concept is that it would make no sense to add evidence that would lower the certainty of estimates—thus, our pooling estimates of higher and lower certainty are based on a belief that the lower certainty estimate is complementing, rather than undermining, the higher certainty estimate. In consequence, since a rating of high certainty is the best one can get, when the direct evidence has high certainty, we can take a shortcut and omit the rating of certainty of the indirect estimate, moving straight to the rating of the network estimate.

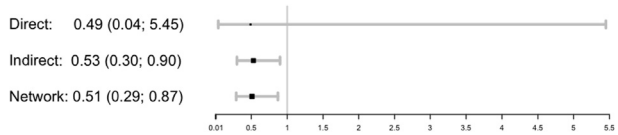

Fig. 2. Direct, indirect, and network estimates for the alendronate vs. raloxifene comparison.

The logic of not rating the certainty of the indirect evidence if the direct evidence is high—without, bear in mind, considering precision—implies, therefore, that reviewers need to ensure that the direct evidence is contributing to the network estimate at least as much as the indirect evidence. This can be done, for instance, by comparing the widths of the confidence intervals of the direct and indirect estimates: the estimate that has the narrower confidence interval is the one contributing to the network estimate the most. We describe this with more details in Appendix 1 at www.jclinepi.com. It is important to note, however, that assessing whether the direct evidence contributes to the network estimate at least as much as the indirect evidence is not the same as examining for imprecision: depending on the precision of the indirect estimate, both a precise and imprecise direct estimates can contribute importantly to a network estimate.

The reason for making sure that the direct evidence contributes to the network estimate at least as much as the indirect evidence is that the certainty rating of the network estimate should be based primarily on the evidence that most contributes to that estimate. It would not, therefore, be appropriate to neglect the certainty of the indirect evidence if the direct evidence is contributing little to the network estimate.

For instance, in a published systematic review, authors assessed the impact of resuscitative fluids on mortality [4]. When rating the certainty of the direct evidence for the comparison of high-molecular-weight hydroxyethyl starch vs. balanced crystalloid, the authors judged that the direct estimate had no limitations in risk of bias, inconsistency, indirectness, or publication bias. Therefore, the certainty of this direct estimate was high. In addition, based on an assessment of the width of the confidence intervals, the direct evidence made a larger contribution to the network estimate more than did the indirect evidence: the direct estimate had a confidence interval from 0.99 to 1.30 and the indirect estimate confidence interval from 0.13 to 5.14. Since the direct estimate had high certainty and was clearly contributing to the network estimate more than the indirect estimate (its confidence interval is far narrower), authors did not need to go to the trouble of rating the indirect estimate to inform the rating of the network estimate.

Note that, although we are suggesting to, in some cases, not rate the indirect evidence, this does not mean that we are ignoring the indirect evidence. We have assessed the coherence between the direct and indirect evidence, their relative contribution to the network estimate, and have pooled the direct and indirect evidence to generate the network estimate.

Another example illustrates that high certainty direct evidence will not always dominate the network estimate and, when it does not, rating the indirect estimate is required. The authors of a systematic review compared the impact of 11 pharmacological agents on the risk of fragility fractures [10]. For the direct comparison between alendronate and raloxifene, there were no concerns of risk of bias, inconsistency, indirectness, or publication bias. Therefore, following the first principle presented in this article—that is, skipping the assessment of imprecision when we are

rating the direct estimate to inform the assessment of the network estimate—the direct evidence had high certainty.

When looking at the direct and indirect estimates of effect, however, it becomes clear that one cannot skip the rating of the indirect evidence. A comparison of the width of these confidence intervals shows that, because it has a far narrower confidence interval, the indirect evidence is dominating the network estimate (Fig. 2), and thus, it would not be appropriate to disregard the certainty of the evidence associated with the indirect estimate.

Considering the relative weight of the direct and indirect estimates is equally important when the direct and indirect estimates are incoherent. In other words, when the direct estimate and the indirect estimate show estimates of effect that are incoherent with one another (which in NMA terminology is also known as local inconsistency), their relative contribution to the network estimate is even more relevant: in the face of incoherence, one needs to base the certainty rating on the evidence that most contributes to the network estimate. When incoherence is present, however, one needs to rate down the network evidence further.

For instance, consider a direct estimate with no serious concerns in any domain except for imprecision because it has a very wide confidence interval. Consider in the same situation that the indirect estimate with much narrower confidence interval is making the dominant contribution to the network estimate. The indirect estimate, however, is rated down for risk of bias and indirectness in one or both of the direct comparisons that contribute to this indirect estimate. In such a situation, one should begin the rating of the network estimate as low and end up as very low because of incoherence.

We have suggested using the width of the confidence intervals to assess which evidence contributes to the network estimate the most. One could use, as an alternative, the contribution matrix. This matrix provides, as a percentage, the extent to which each of the direct comparisons included in the network contribute to the network estimates. Therefore, reviewers can use the matrix to derive the proportion of the information contributing to each network estimate that comes from direct and indirect evidence [11]. Unfortunately, the contribution matrix can only be obtained when using a frequentist framework to conduct an NMA.

Whatever the chosen approach, deciding whether the direct evidence contributes to the network estimate as much as the indirect evidence is a matter of judgment. Reviewers must consider whether 50% is enough or if for some cases, slightly more or slightly less is needed to ensure that the network estimate rating is based on the appropriate evidence. As with every judgment, transparency about the process is fundamental.

The previous two issues in the evolution of GRADE guidance for NMAs help achieve efficiency. The subsequent issues are important but provide little help in enhancing efficiency.

3.3. We should not trust a statistical test of global incoherence of the network to assess incoherence at the pairwise comparison level

The following discussion assumes that NMA authors have considered conduct of a meta-regression to explain heterogeneity and incoherence and that residual incoherence exists following any such analyses.

NMA review authors can assess statistical incoherence between direct and indirect evidence at the paired comparison level (using local tests such as a loop-specific comparison, composite tests, node-splitting, and back calculation) or at the network level (using global tests such as the Lu and Ades model, the design by treatment interaction model, and the Q statistic for inconsistency) [12]. Tests for global incoherence address whether there is sufficient incoherence between the direct and indirect evidence to conclude that incoherence exists in at least one of the loops of the network.

Systematic review authors may, facing nonsignificant tests for global incoherence, infer that there is no incoherence in any of the pairwise comparisons. This represents a tempting short cut—no need to assess incoherence for any pairwise comparison.

Although a potentially timesaving strategy, there are dangers associated with this interpretation. It is possible that, as a nonsignificant test of global incoherence suggests, chance alone explains incoherence looking across an entire network. It is also possible, however, that despite that nonsignificant global test, there is true incoherence for one or more of the key comparisons within the network. Such incoherence, if

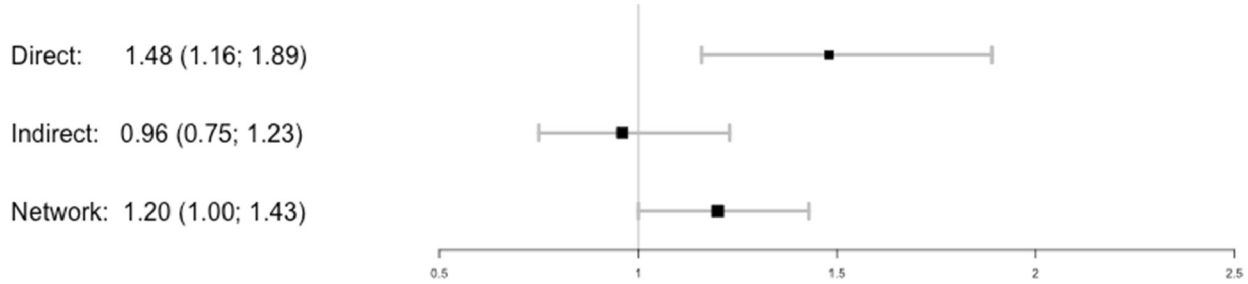

Fig. 3. Qualitative comparison of the direct and indirect estimates of effect.

Table 2. Examples and rational of the conceptual advances to rate the certainty from NMA estimates

Conceptual advance	Example	Direct estimate	Certainty direct estimate	Indirect estimate	Certainty indirect estimate	Network estimate	Certainty network estimate	Guidance	Explanation
Consideration of imprecision is not necessary when rating the direct and indirect estimates to inform the rating of NMA estimates	H2RA vs. sucralfate to prevent stress ulcers in mechanically ventilated critically ill patients. Outcome: pneumonia	1.32 (0.98 to 1.77)	Low (due to risk of bias and imprecision)	1.35 (0.64 to 2.86)	Low (due to risk of bias and imprecision)	1.30 (1.08 to 1.58)	Moderate (due to risk of bias)	Do not consider imprecision of direct and indirect estimates	Network estimate starts as moderate (both direct and indirect estimates are moderate due to risk of bias), and there is no serious concerns of incoherence or imprecision
There is no need to rate the indirect evidence when the certainty of the direct evidence is high and the contribution of the direct evidence to the network estimate is at least as great as that of the indirect evidence	Starch vs. balanced crystalloid (impact of resuscitative fluids on mortality)	1.14 (0.99 to 1.30)	High	0.81 (0.13 to 5.14)	Not needed	1.13 (0.99 to 1.30)	High	Do not consider the certainty of the indirect evidence to rate the network estimate	There is no serious concerns about the direct evidence, and thus, its certainty is high. In addition, the direct evidence contributes much more to the network estimate than the indirect evidence
There is no need to rate the indirect evidence when the certainty of the direct evidence is high and the contribution of the direct evidence to the network estimate is at least as great as that of the indirect evidence	Alendronate vs. raloxifene for preventing fragility fractures	0.49 (0.04 to 5.45)	High (not considering imprecision)	0.53 (0.30 to 0.90)	Moderate (due to risk of bias)	0.51 (0.29 to 0.87)	Moderate (due to risk of bias)	Consider the certainty of the indirect evidence to rate the network estimate—base the network estimate rating in the evidence that contributes the most to it	Although the direct evidence has high certainty, the indirect evidence is contributing much more to the network estimate. Therefore, the rating of the network estimate should be based on the indirect evidence
We should not trust a statistical test of global incoherence of the network to assess incoherence at the pairwise comparison level	Citalopram vs. escitalopram to treat depression	1.48 (1.16 to 1.89)	High	0.96 (0.75 to 1.23)	Very low (due to risk of bias, indirectness, and intransitivity)	1.20 (1.00 to 1.43)	Low (due to incoherence and imprecision)	Do not use a nonstatistically significant test of global incoherence	Although the test for global incoherence suggested that there was no incoherence beyond chance in the network, a comparison of the direct and indirect estimates suggests that there is concerns about incoherence
In presence of incoherence between direct and indirect evidence, the certainty of the evidence of each the estimates can help users decide which estimate to believe.	Citalopram vs. escitalopram to treat depression	1.48 (1.16 to 1.89)	High	0.96 (0.75 to 1.23)	Very low (due to risk of bias, indirectness, and intransitivity)	1.20 (1.00 to 1.43)	Low (due to incoherence and imprecision)	Believe the higher certainty evidence—in other words, use the direct evidence to inform clinical practice	Limitations of the network evidence have to be acknowledged. Using a higher certainty estimate, even if it is not the network estimate, is more appropriate

Abbreviations: NMA, network meta-analysis; H2RA, histamine-2 receptors antagonist.

it exists, raises the issue of whether one should place one's trust in the higher certainty of the direct or indirect estimates or, alternatively, on the network estimate.

However, failure of the global test to detect statistical incoherence when, for one or more of the paired comparisons, statistical incoherence exists may be a consequence of limited power of the global test [13,14]. A recent study reported that the global test failed to detect local level incoherence in 3 out of 40 NMAs that were assessed using both the global test and local level tests [15]. The power of the global test depends on characteristics of the network such as the sample size and number of events, heterogeneity, and the precision of the estimates [14]. Thus, a reviewer who is reassured by a nonsignificant global test of incoherence may risk making misleading high ratings of certainty in network estimates [1].

For example, in an NMA assessing the efficacy of antidepressants [16], the *P* value of the global test for incoherence was 0.30. The authors declared that although there was statistical incoherence detected in 3 out of 70 comparisons, this was compatible with chance, suggesting that users should not be concerned about incoherence in these comparisons. On a closer look, however, there appears to be compelling evidence of incoherence in at least one pairwise comparison, citalopram vs. escitalopram (Fig. 3).

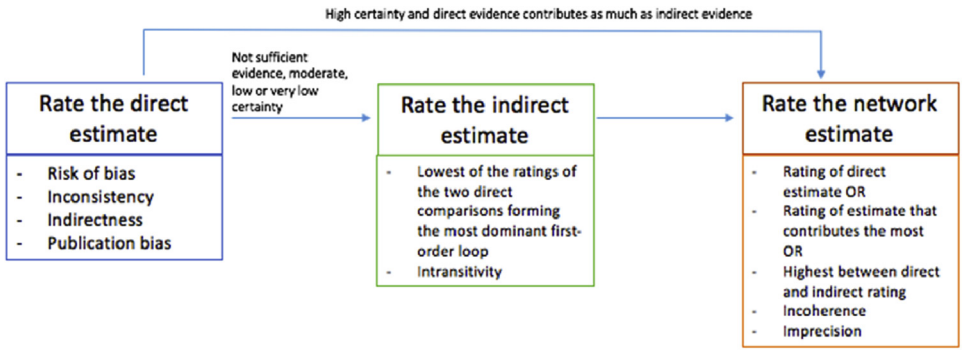

Fig. 4. Process for rating the certainty of the network estimate for each pairwise comparison in an NMA. The figure shows the process to obtain a network estimate rating. Thus, ratings of direct and indirect evidence are illustrated in the context of informing the rating of the network estimate. To obtain a final rating of the certainty of the direct and indirect evidence, imprecision of each of those estimates must be considered.

In assessing incoherence, review authors should consider not only *P*-values, but also the differences in point estimates and the overlap in the confidence intervals. Issues of the network geometry (e.g., presence of multiarm trials) and clinical context may also influence the judgment. In this comparison, the point estimates of the direct and indirect estimates are very different: while the direct estimate suggests a 48% increase in the odds of responding to treatment when patients receive citalopram, the indirect estimate shows a 4% reduction in the odds. In addition, the confidence intervals overlap only minimally, and the Z test comparing these two estimates results in a *P* value of 0.02.

There are further reasons, beyond the statistical tests, to dismiss chance as an explanation for the incoherence. There were serious concerns regarding risk of bias of the direct evidence, and therefore, this was rated as having moderate certainty. The indirect evidence had very low certainty due to risk of bias and indirectness in one of the comparisons forming the loop contributing the most to the indirect estimate and intransitivity between the two comparisons forming this loop. In this instance, we thus have not only statistical evidence of incoherence, but compelling reasons why the indirect estimate is likely to be biased. Systematic reviewers relying on a nonstatistically significant global test for incoherence may miss this incoherence and rate the network estimate higher than is appropriate.

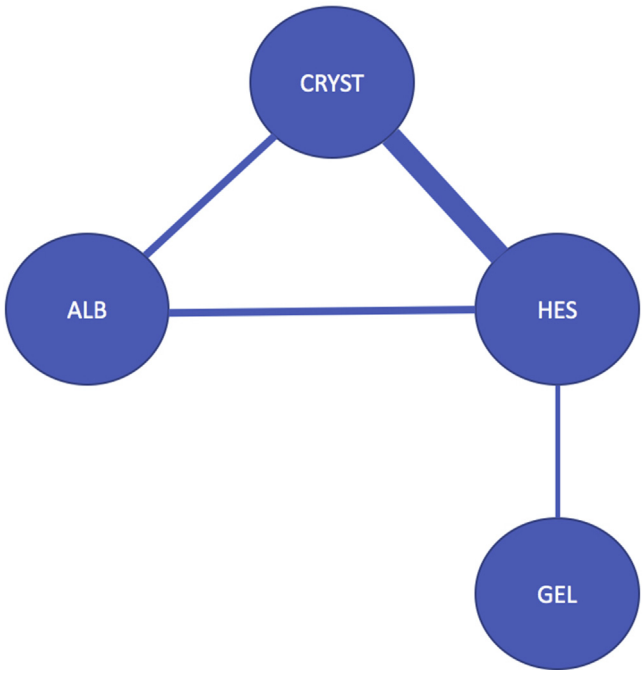

Fig. 5. Network plot of fluid resuscitation in sepsis.

In conclusion, systematic reviewers should ideally assess incoherence in each pairwise comparison. There may be, however, rare situations when review authors reasonably forego the local assessment of incoherence. When dealing with particular networks in which the assessments are very burdensome (well-connected networks with a large number of treatments), reviewers could choose to consider the global test for incoherence and, if it yields a high *P*-value, refrain from considering incoherence at the local level. This would, however, represent a limitation—reviewers may be missing important incoherence at the local level and should acknowledge this limitation in the discussion of their manuscript. When looking at the results of the global test, the lower the *P*-value—even if not meeting usual criteria for statistical significance—the more inclined systematic reviewers should be to refrain from using the shortcut and to assess incoherence for each pairwise comparison.

3.4. In presence of incoherence between direct and indirect evidence, the certainty of the evidence of each estimate can help decide which estimate to believe

Reviewers performing NMA who encounter incoherence in a specific pairwise comparison have two options: (1) rate down the network estimate due to incoherence and use this network estimate acknowledging its limitations or (2) focus on whichever of the direct or the indirect estimates has higher certainty. In the citalopram vs. escitalopram example (Fig. 3), the network estimate would start its rating as moderate (the highest between the direct and indirect rating), but the incoherence would require rating down to low. Moreover, because the confidence interval around the network estimate

Table 3. Results and certainty assessments for the outcome of mortality

Comparison	Direct estimate	Rating	Indirect estimate	Rating	Network estimate	Rating
Starch vs. crystalloid	1.14 (0.99; 1.30)	High	0.81 (0.13; 5.14)	Low[a]	1.13 (0.99; 1.30)	High
Albumin vs. crystalloid	0.81 (0.64; 1.03)	Moderate[b]	1.13 (0.18; 7.32)	Low[a]	0.83 (0.65; 1.04)	Moderate[b]
Gelatin vs. crystalloid	NA	—	1.24 (0.61; 2.55)	Low[a]	1.24 (0.61; 2.55)	Low[a]
Albumin vs. starch	1.40 (0.35; 5.56)	Low[a]	0.71 (0.54; 0.94)	High	0.73 (0.56; 0.95)	High
Gelatin vs. starch	1.09 (0.55; 2.19)	Low[a]	NA	—	1.10 (0.54; 2.22)	Low[a]
Gelatin vs. albumin	NA	—	1.51 (0.71; 3.20)	Low[a]	1.51 (0.71; 3.20)	Low[a]

[a] Very serious imprecision.
[b] Serious imprecision.

includes no effect, rating down for imprecision may also be appropriate. Reviewers and users could focus their attention on a low or very low certainty network estimate or a moderate certainty direct estimate. We would be inclined to believe the moderate certainty direct estimate.

Table 2 summarizes the examples and rationale for each of the conceptual advances we have presented. Based on these advances, we can now think of the GRADE approach for rating the certainty of the evidence from NMA as a three-step process (Fig. 4).

4. Application of these advances to another example

We applied the principles described herein to the NMA assessing the effects of resuscitative fluids on mortality in patients with sepsis [4]. Authors included 14 randomized trials that compared albumin, crystalloid, hydroxyl-ethyl starch, and gelatin to one another (Fig. 5). We used the judgments made by the authors in the original evaluation of the certainty of the evidence for all the GRADE domains. Table 3 presents the results.

Table 4 shows details of the assessments. The only domain for which authors rated down certainty was precision—sometimes rating down one level (serious concern) and sometimes two levels (very serious concern). In terms of efficiency, in every pairwise comparison in which both direct and indirect evidence existed, putting aside concerns of precision, the direct evidence was high quality and contributed to the network estimate as much as the indirect evidence. In consequence, the network estimate was informed only by the direct estimate rating in 3/6 comparisons in this network, and rating of the indirect estimate proved unnecessary. Rating the indirect estimate was necessary only for the comparisons for which there was no direct evidence (gelatin vs. crystalloid and gelatin vs. albumin) and for only one of the comparisons for which there was direct evidence.

Table 4. Details of assessments of certainty of estimates from NMA of resuscitative fluids on mortality in patients with sepsis

Comparison	Starch vs. crystalloid	Albumin vs. crystalloid	Gelatin vs. crystalloid	Albumin vs. starch	Gelatin vs. starch	Gelatin vs. albumin
Direct evidence						
Risk of bias	Not serious	Not serious		Not serious	Not serious	
Inconsistency	Not serious	Not serious		Not serious	Not serious	
Indirectness	Not serious	Not serious		Not serious	Not serious	
Publication bias	Undetected	Undetected		Undetected	Undetected	
Preliminary rating direct	**High**	**High**		**High**	**High**	
Contributes as much as indirect	Yes	Yes		No	Yes	
Need to assess indirect	No	No		Yes	No	
Imprecision	Not serious	Serious		Very serious	Very serious	
Final direct rating	**High**	**Moderate**		**Low**	**Low**	
Indirect evidence						
Common comparator	Albumin	Starch	Starch	Crystalloid		Starch
Tmt1 vs. common comparator rating	High	High	High	High		High
Tmt2 vs. common comparator rating	High	High	High	High		High
Lowest of the two	High	High	High	High		High
Intransitivity	Not serious	Not serious	Not serious	Not serious		Not serious
Preliminary rating indirect	**High**	**High**	**High**	**High**		**High**
Imprecision	Very serious	Very serious	Very serious	Not serious		Very serious
Final indirect rating	**Low**	**Low**	**Low**	**High**		**Low**
Network evidence						
Highest between direct and indirect	High	High	High	High	High	High
Incoherence	Not serious	Not serious	NA	Not serious	NA	NA
Imprecision	Not serious	Serious	Very serious	Not serious	Very serious	Very serious
Final network rating	**High**	**Moderate**	**Low**	**High**	**Low**	**Low**
Most credible estimate	**Network**	**Network**	**Network**	**Network**	**Network**	**Network**

Bold cells represent rating of certainty of direct and indirect estimates, either to inform the network estimate (preliminary ratings) or final ratings. Empty cells indicate that there was no such type of evidence contributing to the network estimate.

The comparison albumin vs. starch provides on illustration from this network showing that addressing imprecision proved necessary only for the network estimate. The 95% confidence interval associated with the direct comparison, 0.35 to 5.56, would require rating down two levels for imprecision. Because the network estimate is dominated by the indirect estimate, it is precise (95% CI: 0.56, 0.95) and mandates a rating of high certainty. Thus, consideration of the precision associated with the direct comparison proved unnecessary.

The incoherence assessment concluded that results were sufficiently similar between direct and indirect comparisons—no serious incoherence existed. Therefore, users can focus on the network estimates to inform clinical practice.

5. Discussion

In this article, we have described and illustrated recent conceptual advances in the GRADE approach for assessing the certainty of the estimates from NMA. The main challenge that reviewers face when rating the certainty of the estimates of effect from NMA is the burden associated with the task. In the original GRADE guidance, for each pairwise comparison, rating a network estimate required an assessment of all five GRADE domains associated with rating down, and for all direct and indirect estimates, sometimes a formidable task. Focusing on the purpose of rating the direct and indirect estimates implicit in the previous guidance—that is, informing the rating of the network estimate—suggested ways to make the process more efficient by foregoing ratings of precision from the direct and indirect estimates and potentially restricting assessment of certainty to the direct estimate when it is high certainty (Table 1). This streamlining of the process results in a reconceptualization of the GRADE approach to certainty of evidence in NMAs as a three-step approach (Fig. 2).

When applying these new concepts, reviewers should bear in mind that although maximizing efficiency is desirable, use of these strategies requires careful judgment. Reviewers should, for instance, even if the direct evidence is high certainty, not skip the assessment of the indirect evidence when the indirect evidence dominates the network estimate. Similarly, inferring from a global test of incoherence that is not statistically significant that there is no incoherence between the direct and indirect estimates of any of the pairwise comparisons is inadvisable: reviewers taking this approach may neglect important incoherence in one or more paired comparisons. Thus, although we are enthusiastic about optimizing the efficiency of the application of the GRADE process for rating quality of evidence in NMAs, we encourage reviewers to complete and report full assessments in their articles. Constructing a table that presents point estimates, confidence or credible intervals, and ratings for all of the direct and network estimates and most of the indirect estimates greatly enhances transparency and usefulness of NMA results.

As we continue using the GRADE approach to rate the certainty of estimates from NMA, we anticipate further developments will arise from the challenges we will encounter. All conceptual advances have been and will be discussed in GRADE working group meetings, and only after full discussion will represent GRADE guidance.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2017.10.005.

References

[1] Puhan MA, Schunemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ 2014;349:g5630.
[2] Balshem 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.
[3] Greco T, Biondi-Zoccai G, Saleh O, Pasin L, Cabrini L, Zangrillo A, et al. The attractiveness of network meta-analysis: a comprehensive systematic and narrative review. Heart Lung Vessel 2015;7(2):133−42.
[4] Rochwerg B, Alhazzani W, Sindi A, Heels-Ansdell D, Thabane L, Fox-Robichaud A, et al. Fluid resuscitation in sepsis: a systematic review and network meta-analysis. Ann Intern Med 2014;161:347−55.
[5] Rochwerg B, Neupane B, Zhang Y, Garcia CC, Raghu G, Richeldi L, et al. Treatment of idiopathic pulmonary fibrosis: a network meta-analysis. BMC Med 2016;14:18.
[6] Sekercioglu N, Angeliki Veroniki A, Thabane L, Busse JW, Akhtar-Danesh N, Iorio A, et al. Effects of different phosphate lowering strategies in patients with CKD on laboratory outcomes: a systematic review and NMA. PLoS One 2017;12(3):e0171028.
[7] Sekercioglu N, Thabane L, Diaz Martinez JP, Nesrallah G, Longo CJ, Busse JW, et al. Comparative effectiveness of phosphate Binders in patients with Chronic Kidney Disease: a systematic review and network meta-analysis. PLoS One 2016;11(6):e0156891.
[8] Rochwerg B, Alhazzani W, Gibson A, Ribic CM, Sindi A, Heels-Ansdell D, et al. Fluid type and the use of renal replacement therapy in sepsis: a systematic review and network meta-analysis. Intensive Care Med 2015;41(9):1561−71.
[9] 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.
[10] Murad MH, Drake MT, Mullan RJ, Mauck KF, Stuart LM, Lane MA, et al. Clinical review. Comparative effectiveness of drug treatments to prevent fragility fractures: a systematic review and network meta-analysis. J Clin Endocrinol Metab 2012;97:1871−80.
[11] Krahn U, Binder H, Konig J. A graphical tool for locating inconsistency in network meta-analyses. BMC Med Res Methodol 2013;13:35.
[12] Efthimiou O, Debray TP, van Valkenhoef G, Trelle S, Panayidou K, Moons KG, et al. GetReal in network meta-analysis: a review of the methodology. Res Synth Methods 2016;7(3):236−63.
[13] Song F, Clark A, Bachmann MO, Maas J. Simulation evaluation of statistical properties of methods for indirect and mixed treatment comparisons. BMC Med Res Methodol 2012;12:138.
[14] Veroniki AA, Mavridis D, Higgins JP, Salanti G. Characteristics of a loop of evidence that affect detection and estimation of inconsistency: a simulation study. BMC Med Res Methodol 2014;14:106.
[15] Veroniki AA, Vasiliadis HS, Higgins JP, Salanti G. Evaluation of inconsistency in networks of interventions. Int J Epidemiol 2013;42:332−45.
[16] Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, Churchill R, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet 2009;373:746−58.