Journal of Clinical Epidemiology 108 (2019) 77−85

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

GRADE approach to rate the certainty from a network meta-analysis: addressing incoherence

Romina Brignardello-Petersen[a,*], Reem A. Mustafa[a,b], Reed A.C. Siemieniuk[a], M. Hassan Murad[c], Thomas Agoritsas[a,d], Ariel Izcovich[e], 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 Street West, Hamilton, Ontario L8S 4L8, Canada*
[b]*Division of Nephrology and Hypertension, Department of Medicine, University of Kansas Medical Center, 3901 Rainbow Blvd, MS3002, Kansas City, KS 66160, USA*
[c]*Mayo Clinic Evidence Based Practice Center, 200 1st Street SW, Rochester, MN 55905, USA*
[d]*Division of General Internal Medicine and Division of Clinical Epidemiology, University Hospitals of Geneva, Switzerland, Rue Gabrielle-Perret-Gentil 4, 1211 Geneve, Switzerland*
[e]*Internal Medicine Service, German Hospital, Pueyrredón 1640, Buenos Aires C1118AAT, Argentina*

Accepted 30 November 2018; Published online 5 December 2018

Abstract

This article presents official guidance from the Grading of Recommendations Assessments, Development, and Evaluation (GRADE) working group on how to address incoherence when assessing the certainty in the evidence from network meta-analysis. Incoherence represents important differences between direct and indirect estimates that contribute to a network estimate. Bias due to limitations in study design or publication bias, indirectness, and intransitivity may be responsible for incoherence. Addressing incoherence requires a judgment regarding the importance of the impact on the network estimate. Reviewers need to be alert to the possibility of misguidedly arriving at excessively low ratings of certainty by rating down for both incoherence and other closely related GRADE domains. This article describes and illustrates each of these issues and provides explicit guidance on how to deal with them.

Keywords: Systematic reviews; Network meta-analysis; Quality of the evidence; Certainty in the evidence; GRADE; Incoherence; Inconsistency

1. Introduction

The Grading of Recommendations Assessments, Development, and Evaluation (GRADE) working group has presented guidance for evaluating the certainty of the evidence (confidence in evidence, quality of evidence) in network meta-analysis (NMA) [1,2]. This article provides additional guidance focused on assessing incoherence and how to address situations in which incoherence may be related to other aspects of certainty judgments. The discussion, which assumes familiarity with the basic concepts of NMA, the concepts related to rating the certainty of the evidence of pairwise comparisons, and the concepts related to rating the certainty of indirect evidence, constitutes official guidance from the GRADE working group. This article was developed and refined by the named authors with feedback from the entire GRADE working group that ultimately approved the article as GRADE guidance.

2. Coherence and incoherence

Coherence, one of the core assumptions of NMA, refers to agreement between direct and indirect evidence. That is, for each pairwise comparison of any two interventions, estimates of the relative effectiveness from the direct and indirect evidence should be similar [3]. For example, if there is both direct evidence to estimate the relative effectiveness of A vs. B and indirect evidence through a common comparator C, the effect estimate obtained using the direct evidence (i.e., the trials that directly compare A vs. B) should be similar to that obtained using the indirect evidence (i.e., the indirect comparison through C).

The GRADE approach for assessing the certainty of evidence from NMA specifies that reviewers must address

Conflict of interest: R.B-P. declares that none of the authors of this article has any conflict of interest to declare.

* Corresponding author. Tel.: +1 9055259140 ext 21635.
E-mail address: brignarr@mcmaster.ca (R. Brignardello-Petersen).

https://doi.org/10.1016/j.jclinepi.2018.11.025
0895-4356/

What is new?

- This is the third article in which the Grading of Recommendations Assessments, Development, and Evaluation (GRADE) working group provides guidance on how to assess the certainty in the evidence from network meta-analysis.
- We present and illustrate guidance on how to address incoherence between direct and indirect evidence, which is a GRADE domain specific to the assessment of evidence from network meta-analysis.
- Rating down the certainty of the evidence due to incoherence requires careful consideration of the impact of this incoherence on the estimates of effect and—to avoid inappropriately rating down twice—establishing the cause of incoherence.

incoherence when finalizing the assessment of each network estimate [2]. For each pairwise comparison resulting from direct and indirect evidence, GRADE directs reviewers to compare the direction and magnitude of the point estimates of the direct and indirect estimates, assess the extent of overlap of the associated confidence intervals (CIs), and consider the results of a statistical comparison of these two estimates. If the direct and indirect estimates are sufficiently dissimilar, reviewers should rate down the certainty of the network estimate for incoherence—and perhaps use the most trustworthy (i.e., the one with the highest certainty of evidence) of the direct or indirect as the best estimate of the relative effect of the paired comparison under consideration.

There are a number of reasons for incoherence: we classify these reasons based on GRADE domains (Fig. 1). First, either the direct or indirect estimates of effect, or both, may be biased due to limitations in the design of the studies or to publication bias (in the case of the indirect evidence, limitations in the study design of studies or publication bias in one or more of the direct comparisons that inform the indirect comparison) (Fig. 1 [1]). Second, either direct or indirect estimates can suffer from indirectness and thus apply to different patients, interventions, or outcomes than the target clinical question (Fig. 1 [2]). Third, intransitivity may result in a biased indirect estimate due to differences—for instance, in the populations enrolled—that modify the effect of the interventions in the direct comparisons informing the indirect comparisons (Fig. 1 [3]). As we will describe below, identifying the cause of incoherence has important implications on whether authors should rate down for incoherence.

Previous GRADE guidance [1,2] has emphasized that rating the certainty of the estimates from NMA requires considering the certainty of both the direct and indirect evidence. The certainty of the network estimate should usually be based on the source of evidence, direct or indirect, that has the largest influence on the network estimate. If both sources of evidence contribute to a similar degree, reviewers should base the network certainty estimate on whichever source has the higher certainty. As the subsequent discussion will make evident, this principle is relevant to the assessment of incoherence. Table 1 summarizes the guidance provided in this article.

When addressing incoherence, reviewers should not rate down for incoherence unless that incoherence is having an important impact on the network estimate. In addition, reviewers should identify the cause of incoherence to avoid rating down due to limitations of the evidence that have already been addressed. Thus, the decision to rate down for incoherence represents a three-step process (Fig. 2).

3. The assessment of incoherence should consider whether both the direct and indirect estimates contribute importantly to the network estimate

As described previously, review authors should address incoherence between the direct and indirect estimates by considering three factors: (1) similarity of point estimates, (2) overlap of CIs, and (3) any statistical test comparing these two estimates. This assessment is analogous to what is done when assessing differences in results between individual studies at the direct comparisons level: in that situation, all the criteria are considered together, using a clinical perspective [4]. If this process yields no concerns, incoherence is not an issue; the assessment stops there, and reviewers would not rate down for incoherence.

If there are concerns, however, reviewers should move to the second step. This step involves determining whether the two estimates are making important contributions to the network estimate (Fig. 2). If, on the one hand, the network estimate is dominated by either the direct or the indirect estimate, the incoherence is impacting minimally on the network estimate, and there is no need to rate down. On the other hand, if both bodies of evidence are influencing the network estimate, rating down (and possibly deciding to believe the more credible of the direct or indirect estimate) is required. The following discussion provides further guidance regarding the second step.

The second step begins with an assessment of the relative contribution of the direct and indirect estimates to the network estimate. One method to assess whether the direct and indirect estimate both contribute importantly to the network estimate is to compare the widths of their CIs. In situations in which the widths of the two CIs are similar, both estimates are making important contributions to the network estimate. In such situations, the network estimate will be intermediate between the direct and indirect estimates, and reviewers should rate down the certainty for incoherence.

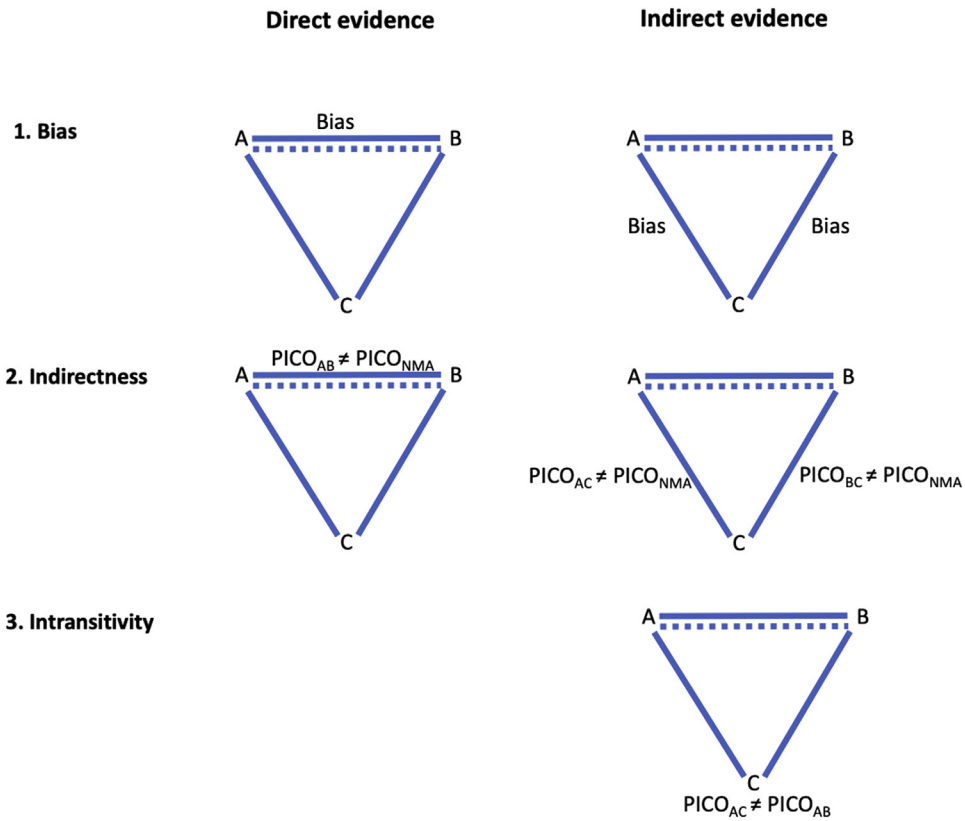

Fig. 1. Causes of incoherence. The comparison of interest is A vs. B. The solid lines represent direct comparisons between the interventions, and the dashed line represents the indirect comparison of A vs. B through the common comparator C. NMA, network meta-analysis; PICO, population, intervention, comparator, outcome.

In other situations, the direct—or indirect—evidence may contribute much more to the network estimate; we will call the estimate that is contributing much more to the network estimate the dominant estimate. When a dominant estimate exists, the associated CI will be much narrower, and the network estimate will be very similar to the dominant estimate (if it is not very similar, then it does not contribute enough to be categorized as dominant). In other words, the nondominant estimate makes only an unimportant contribution to the network estimate. In such situations, reviewers need not rate down for incoherence because the incoherence does not really matter.

Table 1. GRADE guidance for addressing incoherence

The assessment of incoherence should consider whether both the direct and indirect estimates contribute importantly to the network estimate.
The cause of incoherence may be bias in the direct, indirect, or both the estimates or it may be indirectness in the direct, indirect, or both estimates; in either case, reviewers should also rate down the network estimate due to the incoherence.
An additional cause of incoherence may be intransitivity. However, reviewers should seldom rate down a network estimate two levels due to simultaneous intransitivity and incoherence.
Reviewers may appropriately hesitate to rate down a network estimate two levels due to simultaneous incoherence and imprecision because incoherence may be the cause of imprecision.
Reviewers may appropriately rate down a network estimate by two levels due to simultaneous intransitivity, incoherence, and imprecision.
When reviewers rate down the network estimate for incoherence, choosing the highest certainty direct or indirect evidence rather than the network estimate as the best estimate of effect may be preferable.

To illustrate the two steps in deciding to rate down for incoherence, we will use an example from an NMA addressing the impact of alternative surgical approaches to open tibial fractures on reoperation [5]. When comparing unreamed vs. reamed nailing, both the direct and indirect point estimates suggested that unreamed nailing was superior. The magnitude of the effect, however, was much larger in the indirect estimate (odds ratio [OR], 0.07; 95% CI 0.01 to 0.46) than in the direct estimate (OR, 0.74; 95% CI 0.45 to 1.24) (Fig. 3, upper panel). The CIs of the two estimates barely overlapped and the statistical test comparing them resulted in a *P*-value of 0.02. Therefore, the reviewers concluded that there was serious incoherence and rated down the network estimate.

In a prior article, we used this example to illustrate the need to rate down for incoherence [2]. A closer review, however, led to the realization that, despite the differences between the direct and indirect estimates, rating down for incoherence

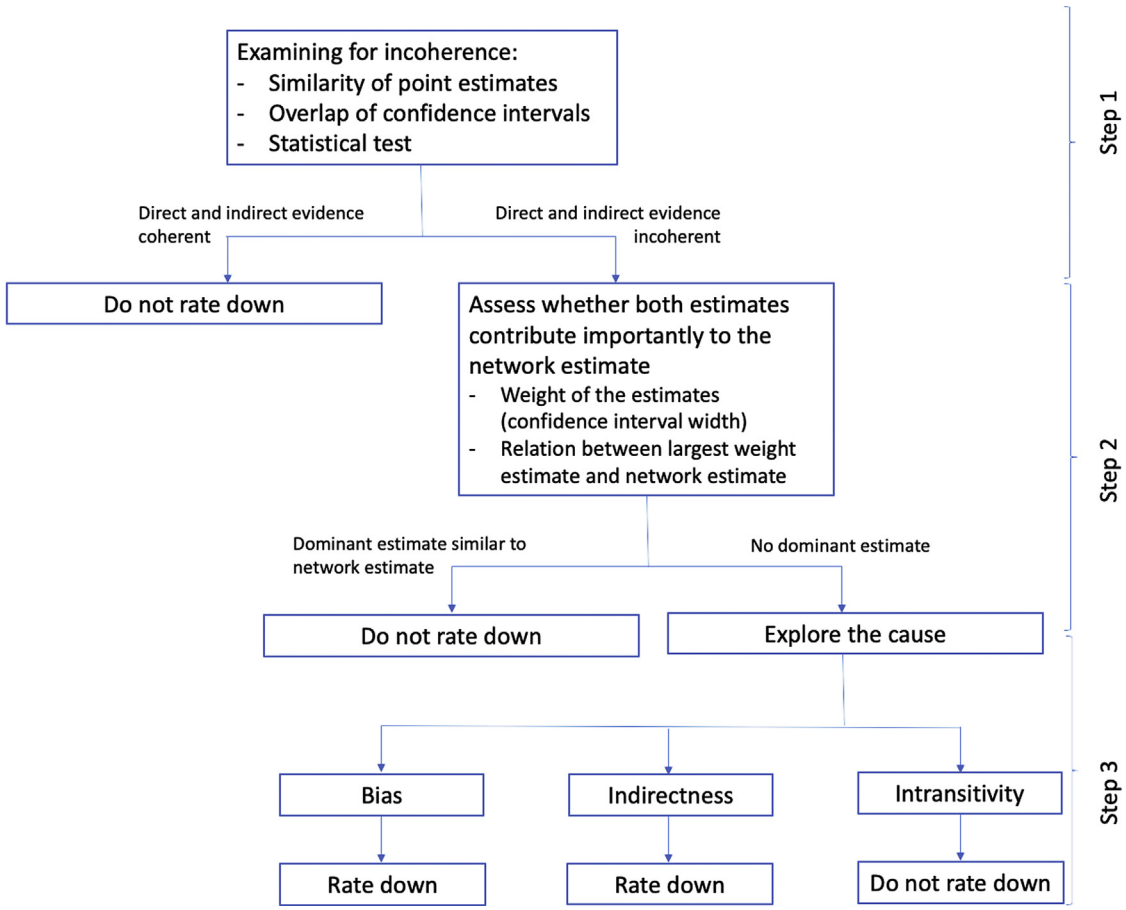

Fig. 2. Rating down for incoherence. All references to rating down are one level. Very rarely reviewers would rate down two levels due to incoherence.

may not be necessary. Here, as shown by the estimates in the log OR scale (Fig. 3, lower panel), the CI around the direct estimate is much narrower than the CI around the indirect estimate. Moreover, the network estimate has a point estimate and CI (OR, 0.62, 95% CI 0.37 to 1.03) very similar to that of the direct estimate (OR 0.74; 95% CI 0.45 to 1.24) (Fig. 3, upper panel). Thus, the direct estimate provides most of the information; the indirect estimate makes an unimportant contribution and does not influence the network estimate sufficiently to rate down for incoherence.

Judging whether there is a dominant estimate requires establishing what it is in which we are rating the certainty [6].

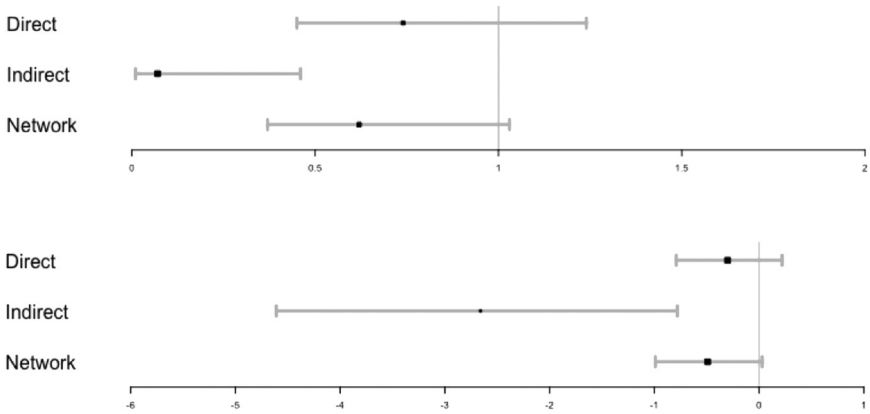

Fig. 3. Direct, indirect, and network estimates comparing unreamed vs. reamed nailing for open tibial fractures. The upper panel shows the estimates and their confidence intervals in the odds ratio scale; the lower panel shows them in the log odd ratios scale. Because a comparison of the widths of the confidence intervals in the odds ratio scale may be misleading due to the shrinking of the scale in one side, authors should make the assessment of the relative contributions of the direct and indirect comparisons in the log odds ratio scale.

Table 2. Assumed risks and absolutes risk reductions of reoperation when comparing reamed vs. unreamed nailing in patients with tibial fracture

Assumed risk for reamed nailing	Direct estimate absolute risk reduction (95% CI)	Network estimate absolute risk reduction 95% (CI)
Low (20 per 1000)	5 less (11 less to 5 more)	7 less (13 less to 1 more)
Moderate (240 per 1000)	60 less (130 less to 60 more)	80 less (150 less to 10 more)
High (400 per 1000)	104 less (220 less to 96 more)	152 less (252 less to 8 more)

Abbreviation: CI, confidence interval.

We show only direct and network estimates because, as previously described, the indirect evidence made a small contribution to the network estimate.

One approach is to rate our certainty regarding the course of action the evidence suggests, that is, would clinical action differ if one uses the estimate that contributes much more to the network estimate vs. the network estimate. This would clearly be appropriate if the NMA was informing a guideline and would then involve considering all relevant outcomes in the specific context of that guideline. This would require considering differences between the estimates in the absolute scale [7] because these are the estimates that clinicians, patients, and guideline panels use to make decisions.

Table 2 illustrates the judgments regarding whether differences between direct and network estimates are important by assuming a low, moderate, and high risk of reoperation with reamed nailing, to judge to what extent the impact of incoherence between direct and indirect estimates is important. If the risk of reoperation with reamed nailing is low, both the direct and network estimates show a similar absolute risk reduction, and most clinicians would not choose one technique over the other based on differences in this outcome. On the other hand, if the risk of reoperation is high, the network estimate shows a much larger absolute risk reduction than the direct estimate; many clinicians may choose unreamed nailing based on the network estimate but not on the direct estimate. Consequently, the impact of incoherence is important and may lead to rating down the certainty in the network estimate if the absolute risk of reoperation is high but not if it is low.

A second approach involves a completely noncontextualized perspective, such as rating certainty in a nonzero effect, in which case, the decision would be straightforward. We can illustrate this third approach using the same example (Fig. 2). Using the noncontextualized approach in which reviewers rate the certainty that there is a nonzero effect, the direct and the network estimates yield the same interpretation because the CI in both cases crosses our threshold OR of 1.0; the evidence does not establish the superiority of unreamed nailing. Therefore, pooling these two incoherent pieces of evidence did not affect our conclusions about the superiority of one treatment vs. the other, and reviewers would not rate down for incoherence. They would, however, rate down the network estimate due to imprecision.

We have established that the second step of assessing incoherence, that is, when incoherence is present, deciding whether to rate down—requires determining whether both the direct and indirect estimates contribute importantly to the network estimate. We will now focus on situations in which this is the case, that is, incoherence exists, and both estimates make important contributions, suggesting the need to rate down the certainty of a network estimate for incoherence (step 3).

4. The cause of incoherence may be bias in the direct, indirect, or both estimates or it may be indirectness in the direct, indirect, or both estimates; in either case, the reviewer should also rate down the network estimate due to the incoherence

The GRADE approach recognizes that, when rating the overall quality of evidence for a particular outcome, reviewers should look at the whole picture rather than a completely separate examination of each of the five reasons for rating down [8]. This includes avoiding double counting limitations in the body of evidence that bear on more than one of the GRADE domains. In other words, reviewers should not rate down 2 levels when there is sufficient overlap in two domains in which they have serious concerns.

Based on this guidance, some may think that whenever reviewers have rated down for bias (due to limitations in study design or publication bias), the evidence, whether direct or indirect, that drives the certainty rating of the network estimate, they should not rate down a second time for incoherence. The logic would be that the bias is counted twice, once in rating down the direct or indirect estimate that provides the basis for the rating of the network estimate and a second time when the reviewer rates down for incoherence.

Although the argument has some appeal, it would lead to problematic practice. Consider a situation in which a direct estimate has serious risk of bias due to limitations in study design (moderate certainty evidence) and an indirect estimate has serious risk of bias and indirectness (low certainty evidence). Consider, also, that these two estimates are incoherent and that both are contributing importantly to the network estimate. In this case, the direct estimate determines the initial rating of certainty (moderate) but by the reasoning, we have laid out previously—important differences between the network estimate and the contributing direct and indirect estimates exist (i.e., incoherence)—reviewers should rate down that estimate to low certainty.

That decision stands despite the rating down for risk of bias. The reason is that, irrespective of the decision to rate

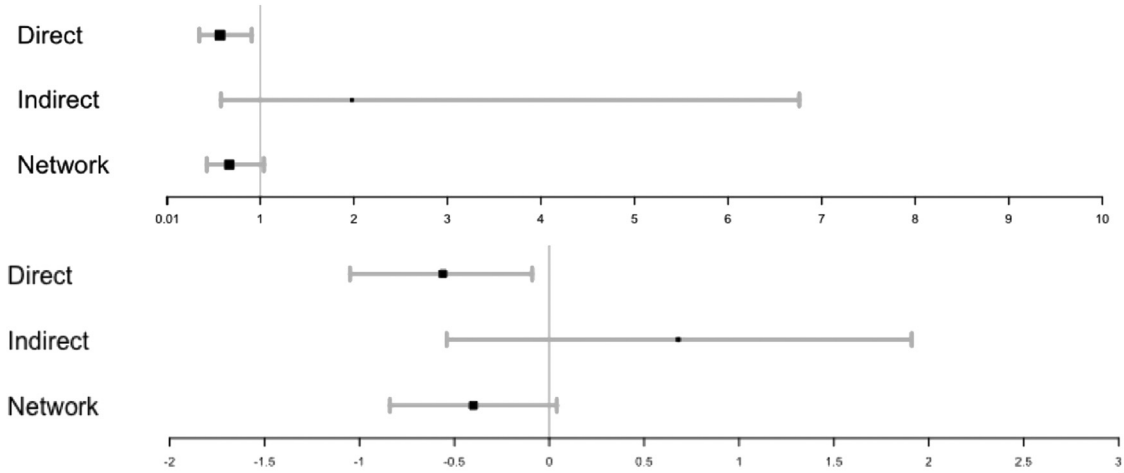

Fig. 4. Effect of placebo vs. misoprostol on gastrointestinal serious adverse events. The upper panel shows the estimates and their confidence intervals in the odds ratio scale; the lower panel shows them in the log odd ratios scale. Because a comparison of the widths of the confidence intervals in the odds ratio scale may be misleading due to the shrinking of the scale in one side, authors should make the assessment of the relative contributions of the direct and indirect comparisons in the log odds ratio scale.

down for bias, our trust in the certainty of this network estimate—which is intermediate between the moderate quality direct estimate and the low quality indirect estimate—would be greater if results of direct and indirect estimates were similar. The additional uncertainty we have in the network estimate needs to be captured in our overall rating of the certainty of the network estimate.

For example, when comparing the effects of nonsteroidal anti-inflammatory drugs on serious gastrointestinal events [9], the comparison between placebo and misoprostol showed a direct estimate of OR 0.57 (95% CI 0.35 to 0.91, moderate certainty due to risk of bias), an indirect estimate of OR 1.98 (95% 0.58 to 6.76, very low certainty due to risk of bias, intransitivity, and imprecision), and a network estimate of OR 0.67 (95% CI 0.43 to 1.04) (Fig. 4). The statistical test comparing these two estimates had a *P* value of 0.06.

In this example, the direct estimate is contributing to the network estimate more than the indirect estimate, and therefore, the certainty of the network estimate should be based on the direct estimate. When examining for incoherence, there are differences in the direction of the point estimates, little overlap of the CIs, and a low *P* value in the test for incoherence; reviewers would therefore appropriately conclude that incoherence exists. Thus, the first step leads to the conclusion that incoherence is present.

Applying the second step, although the direct and network estimates do not differ substantially in their point estimates, they do differ in the precision of their CIs. One would conclude from the direct estimate, with moderate certainty due to bias owing to limitations in study design, that placebo results in at least a 9% relative decrease in gastrointestinal events. This conclusion is not nearly so secure in the network estimate, in which one boundary of the CI suggests a 4% relative increase in events with placebo. Using a completely noncontextualized perspective in which we rate the certainty on a nonzero effect, both estimates make an important contribution to the network estimate and, thus, the network estimate falls between the direct and the indirect estimates. Thus, incoherence not only exists but warrants a rating down of the certainty of the evidence. This example illustrates that regardless of having rated down for risk of bias in the direct estimate, incoherence affected the network estimate sufficiently to further decrease certainty.

This logic also applies to situations in which the cause of incoherence is the indirectness of the direct or indirect estimates, and this indirectness was already accounted for when rating the estimate that contributed the most to the network estimate. In summary, whether the apparent cause of incoherence is bias or indirectness associated with the direct comparisons, if there is serious incoherence, researchers should also rate down for incoherence.

5. An additional cause of incoherence may be intransitivity. However, reviewers should seldom rate down a network estimate two levels due to simultaneous intransitivity and incoherence

Transitivity, which implies that the direct comparisons that contribute to an indirect comparison are similar enough that we need not be overly concerned about effect modification biasing the indirect comparison, is one of the underlying assumptions of NMA [10]. For instance, for an indirect estimate of the relative effect of A vs. B using a common comparator C, the transitivity assumption states that the A vs. C and B vs. C comparisons involve sufficiently similar patients, cointerventions, outcomes, and risk of bias that we do not suspect effect modification, that both A and B have been administered similarly in the direct and

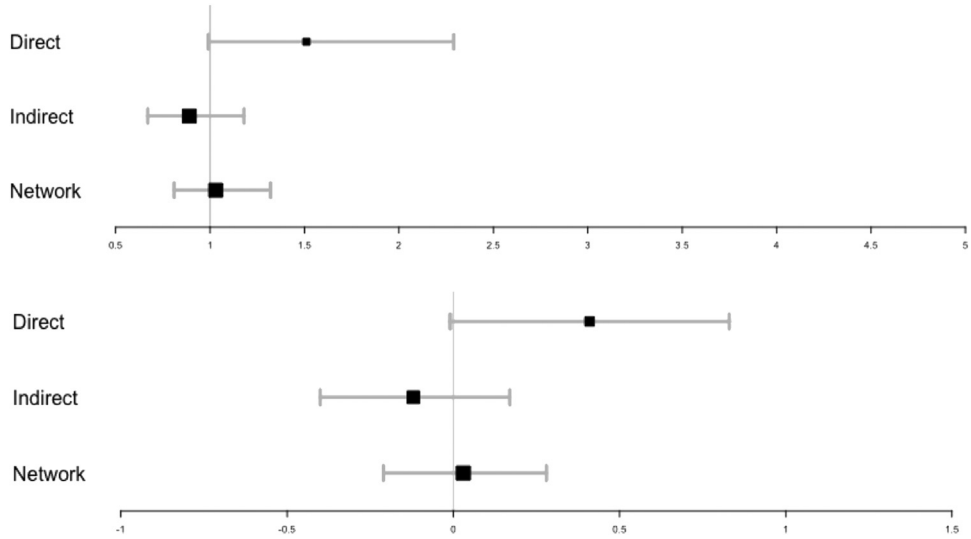

Fig. 5. Direct, indirect, and network estimates comparing venlafaxine vs. mirtazapine. The upper panel shows the estimates and their confidence intervals in the odds ratio scale; the lower panel shows them in the log odds ratio scale. Because a comparison of the widths of the confidence intervals in the odds ratio scale may be misleading due to the shrinking of the scale in one side, authors should make the assessment of the relative contributions of the direct and indirect comparisons in the log odds ratio scale.

indirect comparisons, and the C in A vs. C is sufficiently similar to the C in B vs. C.

Methodologists have explained the transitivity assumption using different interpretations [11]: (1) the characteristics of the two sets of trials are sufficiently similar that there is no important effect modification; (2) the common comparator in an indirect comparison is similar when it appears in both comparisons informing the indirect comparison; (3) the missing treatments in each trial on a loop are missing at random; (4) there are no differences between the relative effects of the direct comparisons forming an indirect comparison beyond what can be explained by heterogeneity; (5) and participants included in the network could have been randomized to any of the treatments. These conceptualizations are nothing more than alternative ways of explaining or understanding the assumption of transitivity. Some audiences may find one conceptualization easier to grasp or more compelling than another, but all conceptualizations provide ways of understanding the same underlying concept: if the condition described is met, the indirect comparison is not biased.

The GRADE approach for assessing the certainty of evidence from NMA establishes that reviewers must address the possibility of intransitivity when assessing the certainty in the indirect evidence contributing to a network estimate [2]. Using the interpretation of intransitivity that, in our experience, clinicians and nonstatistician review authors find easiest to grasp, reviewers should compare patients, interventions, cointerventions, methods and timing for outcome measurement, comparators, and risk of bias between the direct comparisons contributing to an indirect comparison. If these are not similar enough (in other words, if there is an important difference in any potential effect modifier between the direct comparisons forming an indirect comparison), reviewers should rate down the certainty in the indirect evidence for intransitivity (Fig. 1).

Intransitivity and incoherence are closely related. Because the worry with intransitivity is that it will bias indirect estimates of the paired comparison of interest, if the direct and indirect comparisons are similar, one could argue that important intransitivity (i.e., important effect modification in the direct comparisons that inform the indirect comparison) is absent. That logic might lead one to refer to incoherence as the statistical manifestation of intransitivity [10]. Taking the argument a step further, one could argue that if important intransitivity exists, one would expect direct and indirect estimates to differ and thus, if results are coherent, important intransitivity must be absent. Our view is that this last step would lead to excessive dismissal of the possibility of intransitivity, but the logic makes clear that intransitivity may be one of the causes of incoherence.

In keeping with GRADE's goal of making the process for assessing the certainty of the evidence from NMA as simple as possible, reviewers perform the evaluation of intransitivity and incoherence at different stages: intransitivity is assessed when rating the indirect evidence and incoherence when rating the network estimate [1]. Thus, if intransitivity and incoherence coexist, the intransitivity may well be the cause of incoherence, and rating down a network estimate for both intransitivity and incoherence may well be double counting the same problem. Were this the case, rating down for both would result in an inappropriately low certainty rating for the network estimate. Therefore, when the certainty of the network estimate is based on indirect evidence, reviewers should, in most cases, take care not to rate down a network estimate for both incoherence and intransitivity.

To illustrate, consider an NMA assessing the acceptability of antidepressants [12] in which the direct and indirect evidence that contributed to the network estimate were incoherent (Fig. 5), whereas the direct estimate suggests a 51% increase in the odds of accepting the treatment when patients receive venlafaxine, the indirect estimate shows an 11% reduction in the odds. In addition, the CIs have limited overlap, and the *P*-value comparing these two estimates is 0.05.

The figure also shows, from its narrower CI, that the indirect evidence contributes to the network estimate more than the direct evidence. The network estimate certainty rating should therefore be based on the indirect evidence. Reviewers assessed the certainty of the indirect estimate as low due to indirectness in both the direct comparisons that informed the indirect estimate (i.e., both the comparisons between venlafaxine and mirtazapine vs. the common comparator—fluoxetine—had been rated down for indirectness) and intransitivity.

Reviewers rating the network estimate comparing venlafaxine vs. mirtazapine should, on the basis of rating of the indirect comparison certainty as low due to indirectness and intransitivity, start the network estimate rating as low. Because this rating already considers intransitivity concerns, reviewers should not rate down the certainty further due to incoherence; otherwise, they would be double counting this limitation of the network estimate.

6. Reviewers may appropriately hesitate to rate down a network estimate two levels due to simultaneous incoherence and imprecision because incoherence may be the cause of imprecision

Consider once again the example of misoprostol vs. placebo (Fig. 4). Reviewers might be tempted to rate down the network estimate for both incoherence and imprecision, that is, to very low certainty evidence. Because the imprecision has occurred because of incoherence (i.e., combining a precise direct estimate with an incoherent indirect estimate resulted in an imprecise network estimate), this would be unwise. Reviewers could attribute rating down the network estimate, relative to the direct estimate, from moderate to low, to either incoherence or imprecision.

7. Reviewers may appropriately rate down a network estimate by two levels due to simultaneous intransitivity, incoherence, and imprecision

In the example of venlafaxine vs. mirtazapine (Fig. 5), we made the case for not rating down twice due to simultaneous intransitivity and incoherence because intransitivity was the cause of incoherence. In the same example, however, incoherence resulted in serious imprecision. In the previous section, we suggested that rating down the certainty of a network estimate twice due to incoherence and imprecision may not be always appropriate. In the venlafaxine vs. mirtazapine example, intransitivity, incoherence, and imprecision are related. It may be appropriate to rate down two levels owing to these three serious concerns.

8. When reviewers rate down the network estimate for incoherence, choosing the highest certainty direct or indirect evidence rather than the network estimate as the best estimate of effect may be preferable

The issue of what source of evidence constitutes the best estimate applies to all our examples in which rating down the network estimate for incoherence is deemed appropriate. One could ask, in our view very legitimately, why one would choose as the best estimate lower certainty evidence when higher certainty evidence is available. For instance, in the misoprostol example (Fig. 3), the reviewer has moderate certainty evidence from the direct estimate and low certainty evidence from the network estimate. In general, GRADE guidance would suggest choosing the higher certainty direct estimate rather than the network as the best estimate of the paired comparison effect. This does not mean, however, that anything in the NMA or assessment of the certainty of evidence will change. Rather, the NMA authors and those using its results would consider the estimate with the higher certainty, rather than the network estimate, as the more trustworthy estimate.

Although some may argue that picking and choosing between network and direct or indirect estimates is inappropriate and that logical coherence dictates consistent use of NMA estimates as best evidence, we believe that clinical decision-making should be informed by estimates in which we have highest certainty. In all or almost all circumstances, NMA estimates will be, for most paired comparisons, the most trustworthy of the available estimates.

9. Conclusion

Addressing the impact of incoherence in the certainty of evidence from a network estimate and its associated CI requires not only considerations of similarity between direct and indirect estimates of effect but also an assessment of the contribution of the two sources of evidence and a judgment of the extent to which incoherence is serious enough to warrant rating down. In addition, to avoid double counting and making spuriously low ratings of the certainty of network estimates, reviewers must be aware of the relationship between incoherence and other potential limitations in the body of evidence.

References

[1] Brignardello-Petersen R, Bonner A, Alexander PE, Siemieniuk RA, Furukawa TA, Rochwerg B, et al. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. J Clin Epidemiol 2017;93:36−44.

[2] 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.
[3] Lu G, Ades A. Assessing evidence inconsistency in mixed treatment comparisons. J Am Stat Soc 2006;101:447−59.
[4] 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.
[5] Foote CJ, Guyatt GH, Vignesh KN, Mundi R, Chaudhry H, Heels-Ansdell D, et al. Which surgical treatment for open tibial shaft fractures results in the fewest reoperations? A network meta-analysis. Clin Orthop Relat Res 2015;473:2179−92.
[6] Hultcrantz M, Rind D, Akl EA, Treweek S, Mustafa RA, Iorio A, et al. The GRADE Working Group clarifies the construct of certainty of evidence. J Clin Epidemiol 2017;87:4−13.
[7] 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.
[8] Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol 2011;64:1311−6.
[9] Brown TJ, Hooper L, Elliott RA, Payne K, Webb R, Roberts C, et al. A comparison of the cost-effectiveness of five strategies for the prevention of non-steroidal anti-inflammatory drug-induced gastrointestinal toxicity: a systematic review with economic modelling. Health Technol Assess 2006;10:iii−iv. xi−xiii, 1−183.
[10] 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;7:236−63.
[11] Salanti G. Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. Res Synth Methods 2012;3:80−97.
[12] 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.