Journal of Clinical Epidemiology 105 (2019) 60–67

Journal of Clinical Epidemiology

ORIGINAL ARTICLE

GRADE approach to rate the certainty from a network meta-analysis: avoiding spurious judgments of imprecision in sparse networks

Romina Brignardello-Petersen[a], M. Hassan Murad[b,*], Stephen D. Walter[a], Shelley McLeod[a,c], Alonso Carrasco-Labra[a,d], Bram Rochwerg[a,e], Holger J. Schünemann[a], George Tomlinson[f,g], 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, ON L8S 48L, Canada*
[b]*Evidence-Based Practice Center, Mayo Clinic, 200 1st Street SW, Rochester, MN 55905, USA*
[c]*Department of Family and Community Medicine, Schwartz/Reisman Emergency Medicine Institute, University of Toronto, 200 Elizabeth Street, Toronto, ON M5G 2C4, Canada*
[d]*Evidence-Based Dentistry Unit, Faculty of Dentistry, Universidad de Chile, 200 1st Street SW, Rochester, MN 55905, USA*
[e]*Department of Medicine, McMaster University, 1280 Main St W, Hamilton, ON L8S 48L, Canada*
[f]*Department of Medicine, UHN and Mt Sinai Hospital, 200 Elizabeth Street, Toronto, ON M5G 2C4, Canada*
[g]*Institute of Health Policy, Management and Evaluation, University of Toronto, 4th Floor, 155 College St, Toronto, ON M5T 3M6, Canada*

Accepted 17 August 2018; Published online 22 September 2018

Abstract

When direct and indirect estimates of treatment effects are coherent, network meta-analysis (NMA) estimates should have increased precision (narrower confidence or credible intervals compared with relying on direct estimates alone), a benefit of NMA. We have, however, observed cases of sparse networks in which combining direct and indirect estimates results in marked widening of the confidence intervals. In many cases, the assumption of common between-study heterogeneity across the network seems to be responsible for this counterintuitive result. Although the assumption of common between-study heterogeneity across paired comparisons may, in many cases, not be appropriate, it is required to ensure the feasibility of estimating NMA treatment effects. This is especially the case in sparse networks, in which data are insufficient to reliably estimate different variances across the network. The result, however, may be spuriously wide confidence intervals for some of the comparisons in the network (and, in the Grading of Recommendations Assessment, Development, and Evaluation approach, inappropriately low ratings of the certainty of the evidence through rating down for serious imprecision). Systematic reviewers should be aware of the problem and plan sensitivity analyses that produce intuitively sensible confidence intervals. These sensitivity analyses may include using informative priors for the between-study heterogeneity parameter in the Bayesian framework and the use of fixed effects models.

Keywords: Network meta-analysis; meta-analysis; GRADE; certainty; imprecision; clinical practice guidelines; quality of evidence; evidence-based medicine

1. Introduction

Network meta-analysis (NMA) is becoming increasingly popular [1]. One of the most frequently cited advantages of NMA is that, by combining direct and indirect evidence, the network estimate for the relative effectiveness of two interventions will yield a narrower confidence interval (CI) or credible interval than if reviewers were to rely on direct evidence alone [2–5]. We have seen, however, a number of cases of sparse networks in which, despite the direct and indirect evidence being coherent (i.e., showing similar effects), the CIs or credible intervals of the network estimates are considerably wider than those of the direct estimates. In this article, we describe the cause of this phenomenon and provide guidance on how to avoid making spurious assessments of the certainty of the evidence of network estimates in NMAs in which it occurs. The discussion, which assumes familiarity with the basic concepts of NMA, and the concepts related to rating the certainty of the evidence—particularly imprecision—constitute official guidance from the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) working group.

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. 507-284-2560; fax: 507-284-4251.
E-mail address: Murad.mohammad@mayo.edu (M.H. Murad).

https://doi.org/10.1016/j.jclinepi.2018.08.022
0895-4356/

What is new?

Key findings

- When conducting network meta-analysis (NMA) of sparse networks, the common between-study heterogeneity assumption—necessary to conduct the analyses in such networks—may result in network estimates with implausibly wide confidence intervals after combining coherent direct and indirect evidence.

What this adds to what was known?

- There are many statistical methods that can deal with this issue; and although these require making assumptions about the belief and uncertainty on the common between study heterogeneity, they provide sensible results.

What is the implication and what should change now?

- Systematic reviewers who are not aware of this issue or fail to address it will report network estimates that are spuriously imprecise because they result from limitations in the method of statistical analysis and not in the evidence. Systematic reviewers planning to conduct an NMA of sparse networks should plan to conduct sensitivity analyses to address the issue and obtain network estimates more likely to be useful for decision-making.

2. The case of antiarrhythmics for out-of-hospital cardiac arrest

Investigators conducted an NMA comparing the relative effectiveness of five antiarrhythmic drugs and placebo in patients experiencing out-of-hospital cardiac arrest [6]. For the outcome survival to hospital admission, the NMA included eight randomized clinical trials that had enrolled a total of 3,483 participants. The direct estimate for the comparison between lidocaine and placebo for the outcome survival to hospital admission showed a risk ratio (RR) of 1.19 with a 95% CI of 1.07−1.31, suggesting that patients who received lidocaine were 19% more likely to survive to hospital admission than those who received placebo. Because there were no serious concerns about the risk of bias, inconsistency, imprecision, indirectness, or publication bias, the review authors rated the evidence as high certainty.

Although the indirect evidence provided a much less precise estimate, the RR was 1.03 with a 95% CI of 0.53−2.01, and the direct and indirect estimates were coherent (i.e., small differences in the point estimates, with widely overlapping CIs). However, when combining these two sources of evidence using a frequentist framework and a random effects model, the CI of the network estimate was wider than one would anticipate, with a lower bound of 0.85 and an upper bound of 1.45 (Fig. 1); here, instead of increasing the precision of the direct estimate, the NMA decreased precision.

3. NMA assumptions

In addition to transitivity and coherence—the core assumptions of NMA—there is an assumption in the statistical methods for NMA available in commonly used software packages that the between-study heterogeneity is the same across all the comparisons in the network [7,8]. Although this assumption is unlikely to be correct [9], it simplifies the estimation of the NMA parameters [10], and it may be particularly necessary in the case of sparse networks. Using the common heterogeneity assumption, the statistical approaches use information available in the entire network to estimate a common between-study heterogeneity, which is then included in the calculations for all CIs for effect estimates in the network.

In the NMA of antiarrhythmics for out-of-hospital cardiac arrest, the direct estimate of survival to hospital admission for lidocaine versus placebo came from one large study with 2,052 patients. As no estimate of between-study heterogeneity could be calculated, the between-study heterogeneity was assumed to be 0, and the CI for the direct estimate reflected only the within-study variability (this is discussed with more details later in this article). The other comparisons for which there was direct evidence had an estimated between-study heterogeneity up to I^2 = 86% (tau^2 = 0.20). The common between-study heterogeneity estimated for the whole network was I^2 = 48% (tau^2 = 0.02). Therefore, the between-study heterogeneity from the other comparisons affected the precision of the

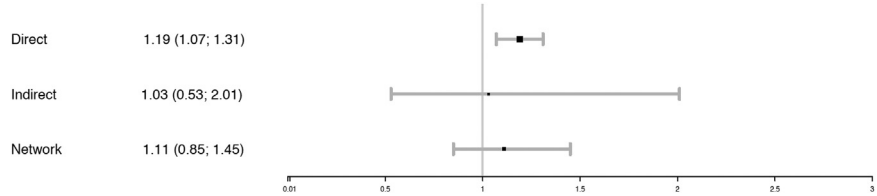

Fig. 1. Direct, indirect, and network estimates of lidocaine versus placebo when using a frequentist random effects model.

NMA estimate for the comparison of lidocaine versus placebo, widening of the CI.

This example illustrates a problem that originates from the common between-study heterogeneity assumption: if, as is likely to be the case, this assumption is not true, some of the network estimates will have CIs that, because of attribution of an inappropriately large between-study heterogeneity, are misleadingly wide. One consequence of this is that reviewers will rate the GRADE certainty of evidence [11,12] down due to concerns of imprecision, and this rating down will be due to a spurious analysis assumption rather than true limitations in the body of evidence. In other words, reviewers would use a poorly estimated heterogeneity parameter that widens the CI of some network estimates dramatically and implausibly, resulting in spurious lower certainty evidence. In this example, combining a high-certainty body of direct evidence comparing lidocaine versus placebo with coherent indirect evidence that carried little information resulted in a spuriously low certainty of evidence rating. The low rating was due to imprecision which was, in turn, due to an almost certainly incorrect assumption of common between-study heterogeneity across comparisons.

Because of the difficulty that many users without programming skills would have and, more importantly, the lack of sufficient data to estimate multiple between-study heterogeneity parameters, methods for allowing the between-study heterogeneity to vary across the network [13] are not feasible in sparse NMAs. Thus, allowing the between-study heterogeneity to vary across the network cannot address the problems that follow from assuming a common between-study heterogeneity in sparse networks. Because results from a survey of 456 published NMAs [1,14] show that at least 25% of the NMAs include fewer than five treatment nodes and fewer than 13 trials, and a meta-epidemiological study of 112 NMAs [15] showed that 36% of statistically significant direct estimates were no longer statistically significant when combined with indirect evidence, this issue may arise frequently, and reviewers need to address it appropriately.

4. Strategies to address the issue

Commonly used statistical software packages provide many statistical methods to conduct NMA. As we have noted, however, most of these methods assume a common between-study heterogeneity across the network. Systematic reviewers must choose both an appropriate model and framework to obtain sensible results when conducting NMA of sparse networks.

4.1. Fixed versus random effects models

The assumptions of the fixed versus random effects models for NMA are the same as those of a traditional meta-analysis. Fixed effects models assume that any observed differences in the relative effects of the interventions between studies are due to chance. In other words, there is a single true treatment effect across all studies [16]. The assumption that all the studies are estimating the same parameter implies that the true between-study heterogeneity is zero, and thus, it has no influence in the estimation of the pooled estimates. As a result, combining data from two studies will always result in a CI that is, because of increased sample size, narrower than that of either study alone.

The assumption behind random effects models is that all studies are estimating related, but not identical, treatment effects. In other words, there is an average treatment effect, and, across studies, there is variability of the true effects around this average [16]. This variation of the treatment effect, the between-study heterogeneity (tau^2), is estimated using the data from all studies and is used to calculate the CI of the pooled estimate. Therefore, the CIs of pooled estimates calculated using random effects models consider both within- and between-study heterogeneities.

Because the between-study heterogeneity can vary, so can the magnitude of the differences between pooled estimates obtained using a fixed versus a random effects model. For the same set of studies, the larger the between-study heterogeneity, the wider the CI of the pooled estimate obtained using the random effects model when compared with that obtained using the fixed effects model.

Because systematic reviews with NMAs tend to address broad rather than narrow clinical questions, it is unlikely that the data will meet the assumptions of a fixed effects model. However, reviewers conducting NMAs who believe that the common between-study heterogeneity across comparisons is unrealistic, or that it cannot be estimated reliably in their sparse networks—and that it is causing some network estimates to have CIs much wider than appears sensible—may reasonably assume that such between-study heterogeneity across comparisons is zero by conducting the NMA using fixed rather than random effects models—if, that is, results make more intuitive sense than those of random effect models.

4.2. Bayesian versus frequentist framework

The Bayesian framework has been most commonly used to conduct NMA [14] probably because software for implementation was available in the early days of NMA [17]. The Bayesian framework combines a prior probability distribution with a likelihood based on the available data, which results in a posterior probability distribution from which the analyst obtains the relative effect estimates [3,18]. To conduct an analysis that is mainly data-driven and because there is often little information about treatments effects outside the data in the meta-analysis, analysts often specify noninformative or weakly informative priors for treatment effects [13].

Noninformative or weakly informative priors contain little to no information about the parameters. Their use indicates a belief that there is no information about the true value of the parameters in the NMA model beyond what the data are indicating and may result in less precise estimates of treatment effects than one would be obtained from frequentist frameworks [19]. This difference in precision of treatment effects arises because, in the Bayesian framework, there can be considerable remaining uncertainty about the value of the between-study heterogeneity parameter with the use of a noninformative or weakly informative prior in a small or sparse network. In the Bayesian framework, this uncertainty propagates through to the credible intervals around estimates, whereas in the traditional frequentist framework, uncertainty around the heterogeneity parameter is usually ignored in the calculation of CIs. Thus, a potential solution to this issue would be to use more informative priors for the between-study heterogeneity, for example, those based on empirical distributions that depend on the type of outcome being investigated and interventions being compared [9]. Analysts can implement such informative priors using the available software packages for conducting Bayesian NMA [10].

Although the frequentist framework is most commonly used for traditional meta-analysis, software allowing for its implementation in NMA by nonstatisticians was developed later than those for Bayesian approaches [14]. The frequentist framework has advantages, including the perceived familiarity of the users with the interpretation of the estimates, as well as making estimates and inferences based entirely on the data available, thus avoiding issues with either informative or uninformative priors.

Considering all the above, reviewers challenged by spurious imprecision of results of NMAs of sparse networks conducted using a Bayesian framework with uninformative or weakly informative priors should adopt one of two alternatives: (1) use informative priors that would place realistic restrictions on the between-study heterogeneity or (2) use a frequentist framework. The choice of a fixed effects approach in a Bayesian setting can be seen as the choice to use an extremely informative prior—one that puts all its weight on a value of zero for the between-study heterogeneity parameter.

5. Planning the statistical analysis of sparse networks

Reviewers will, ideally, plan the statistical methods for pooling studies before the start of their systematic review [20–23]. Reviewers planning to undertake NMA who anticipate that they will have a sparse network must be aware of the potential problem highlighted in this article and, in their protocol, include strategies to deal with the issue should it arise. Considering all the above, reviewers conducting NMA of sparse networks could choose among the following statistical methods that are feasible to implement:

- Bayesian random effects models with uninformative priors
- Bayesian random effects models with informative priors
- Bayesian fixed effects model
- Frequentist random effects models
- Frequentist fixed effects models

It is not possible to know whether the common between-study heterogeneity assumption will be inappropriate before the analysis is undertaken. Reviewers can, however, plan to conduct sensitivity analyses if the common between-study heterogeneity assumption results in excessively imprecise network estimates when using their primary method of analysis. Even though it is unlikely that there is enough data to reliably assess whether the common between-study heterogeneity assumption was appropriate—had that been the case, reviewers could have implemented a model in which the heterogeneity was allowed to vary across comparisons—a close look at each of the pairwise comparisons and their between-study heterogeneity can provide some insight regarding the extent to which the assumption of common between-study heterogeneity across the network was appropriate. If reviewers judge that the common heterogeneity assumption was likely to be inappropriate, they can use alternative statistical frameworks and models that do not rely so heavily on the assumption. Ideally, their protocol will specify a hierarchy of analytic approaches that they will follow only until they arrive at an approach that produces a sensible result.

Finally, although statistical estimates of goodness of fit of the models may seem a suitable approach for determining which model is more likely to provide more appropriate results, believing their results implies that reviewers believe in all the assumptions of the model, including the common between-study heterogeneity across the network. In addition, there is no goodness of fit measure that allows comparing analyses conducted with the Bayesian versus the frequentist framework, decreasing the usefulness of these measures in this context.

6. Comparing the result of alternative approaches

Alternative statistical approaches had an important impact in the network estimate for the relative effectiveness of lidocaine versus placebo for survival to hospital admission in patients experiencing out-of-hospital cardiac arrest (Fig. 2). Even though the indirect evidence provided little information, the NMA assumption of equal between-study heterogeneity across comparisons resulted in network estimates with very wide CI when using the random effects model. As described previously, the reviewers judged that the CI of the network estimate was, as a result of the inappropriate assumption of common between-study heterogeneity across the network, spuriously wide. They, therefore, considered more credible the results from fixed effects models that generated a network CI similar to that of the direct estimate.

The implications of the inference that the NMA estimates using random effects models are inappropriate in this case

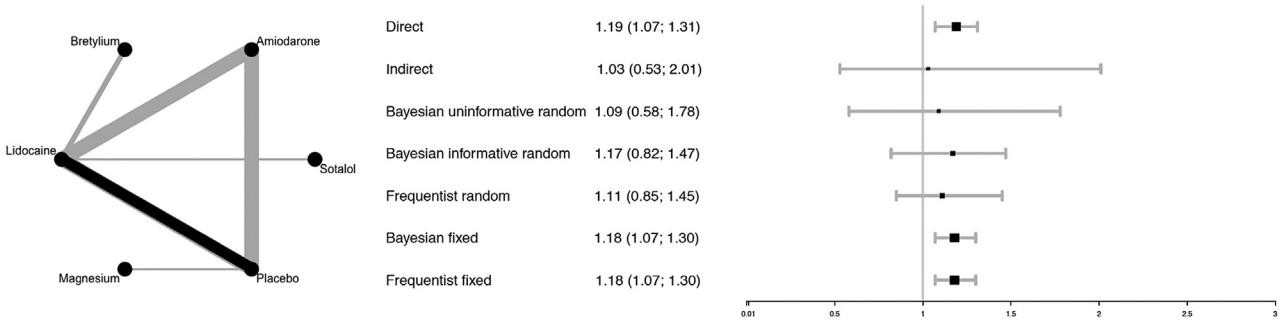

Fig. 2. Network plot and network estimate for the relative risk of the effect of lidocaine versus placebo on survival to hospital admission in patients experiencing out-of-hospital cardiac arrest using different statistical approaches. Bayesian analyses were done using the package *gemtc* [24] in the statistical software R [25]. Bayesian uninformative prior used was the default in the statistical analysis package (*gemtc*, R) [24], and Bayesian informative prior used was taken from a set of empirical priors [9], according to the comparison and outcome. Frequentist analyses were done using the package *netmeta* [26] in R.

are important: reviewers who considered such estimates credible would rate certainty as low owing to very serious imprecision, which would make the evidence less useful for decision-making. The authors of the NMA, however, chose a frequentist fixed effects model that avoided obtaining an implausibly wide CI and an inference regarding certainty of the effect of lidocaine versus placebo that would have resulted from an inappropriate statistical approach.

Application of various statistical methods to other examples shows similar results (Fig. 3). Consider an NMA comparing different agents for dental caries arrest in which there were six treatments compared in seven randomized clinical trials [27]. In a preliminary analysis, for the comparison between fluoride varnish plus toothpaste versus standard oral hygiene, the indirect evidence contributed almost no information to the network estimate. Nevertheless, the CI authors obtained when using a frequentist random effects model was much wider than that of the direct estimate and wide enough to warrant rating down the certainty of the evidence by two levels (Fig. 3, top panel).

Because most of the evidence came from the direct comparison, the reviewers based their conclusions regarding relative effectiveness on the results from a direct estimate in which they had moderate certainty, rather than on a network estimate in which they had very low certainty. As the Figure depicts, conducting the analysis using fixed

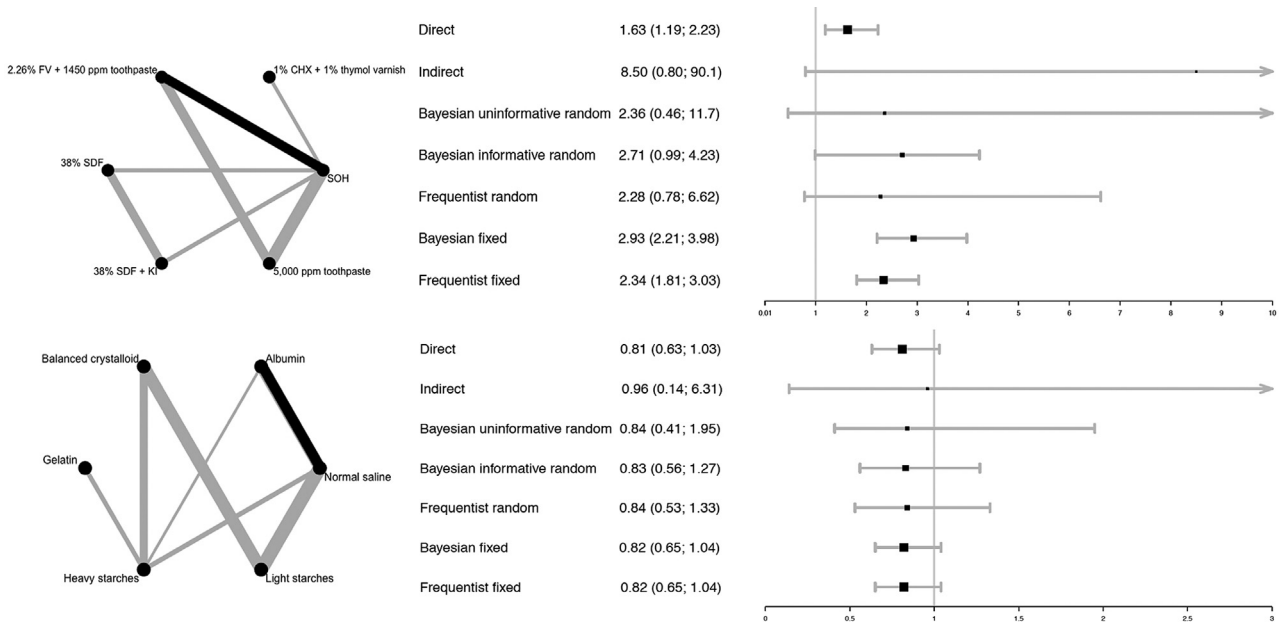

Fig. 3. Top: Network plot and network estimates of relative risk comparing 2.26% fluoride varnish plus toothpaste for caries arrest versus standard oral hygiene. Bottom: Network plot and network estimates of odds ratios comparing the effect of albumin versus normal saline on mortality in patients with sepsis. The comparison of interest is bolded in each network plot. Abbreviations: FV: fluoride varnish; ppm: parts per million; CHX: chlorhexidine; SOH: standard oral hygiene; SDF: silver diamine fluoride; KI: potassium iodide. Bayesian analyses were done using the package *gemtc* [24] in the statistical software R [25]. Bayesian uninformative prior used was the default in the statistical analysis package (*gemtc*, R) [24], and Bayesian informative prior used was taken from a set of empirical priors [9], according to the comparison and outcome. Frequentist analyses were done using the package *netmeta* [26] in R.

effects models would have resulted in a CI similar to that of the direct estimate and thus a rating of the network estimate as moderate certainty evidence.

A similar phenomenon occurred when assessing the effects of intravenous albumin versus normal saline on mortality in patients with sepsis [28], in a network with six treatments and 13 randomized clinical trials (Fig. 3, bottom panel).

Even though these examples illustrate that the choice of statistical approach has an important impact on the network estimates, this impact seems to be specific to a particular comparison within a particular NMA. Although whenever appreciable heterogeneity exists, fixed effects models will result in narrower CIs than random effects models, the magnitude of the narrowing will be much larger in some cases than in others.

7. Our suggested approach is not cherry-picking: NMA protocols should deal with this issue

Although we believe that the correct approach to deal with this problem is to conduct sensitivity analyses that would provide optimally trustworthy network estimates when the assumption of common heterogeneity appears not to be met, some may see this approach as ''cherry picking'' (i.e., conducting multiple analyses and then, on then choosing the approach that yields results they considerable desirable). We argue that this is not the case.

As we have noted earlier in this discussion, NMA authors should, as part of their analysis plan, check results to judge to what extent the assumption of common between-study heterogeneity across the network is credible. As mentioned previously, it is unlikely that there is enough data to check this reliably, and thus the judgment would be based on a detailed look at the data available (i.e, the between-study heterogeneity in each of the network comparisons).

Our examples all included direct evidence that reviewers had judged as high or moderate certainty evidence and coherence between direct and indirect evidence which contributed little additional evidence. We reasoned that adding small additional coherent evidence to high or moderate quality evidence should not result in new estimates of low or very low certainty. Therefore, reviewers should specify, in their protocol, that if they observe noncredible wide CIs in one or more comparisons, they will conclude that the assumption of common between-study heterogeneity across the network is untenable and therefore will select an analysis approach that deals with the problem. That alternative analysis will then become the primary analysis, and review authors should describe the decision-making process and their rationale in the publication of their systematic review. In proceeding in this way, they will ultimately arrive at a presentation of the most trustworthy results to guide clinical care.

8. Unwarranted common heterogeneity assumptions may not be the only reason for counter-intuitively wide CIs/certainty intervals

Consider a systematic review and NMA comparing pharmacologic interventions for patients with nonalcoholic steatohepatitis in which the network was composed of five treatments and seven randomized clinical trials (Fig. 4). In one analysis, the Bayesian random effects model with uninformative priors, combining consistent direct and indirect estimates resulted in an NMA estimate with counterintuitive credible intervals much wider than the direct estimate. This was not true for any of the other analytic approaches, including Bayesian fixed effects models, frequentist random effects models, and Bayesian random effects models with informative priors.

Thus, assumed common heterogeneity does not, in this case, appear to explain the single excessively wide certainty interval. Indeed, in this case, the point estimate of the between-study heterogeneity of both the direct comparison of thiazolidinediones versus placebo and the whole network was 0. Here, the wide credible interval results from the use of an uninformative prior for the heterogeneity parameter in combination with limited data for estimation of this

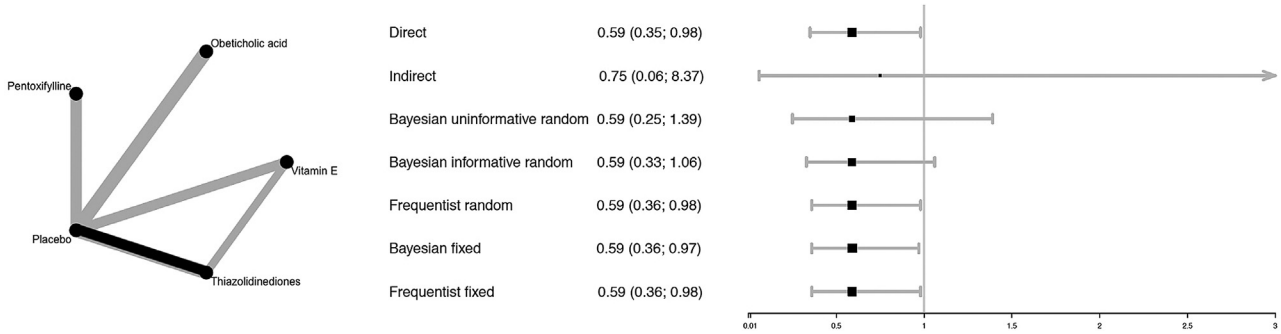

Fig. 4. Network plot and estimates of odds ratio comparing the effects of thiazolidinediones versus placebo on liver fibrosis in patients with nonalcoholic steatohepatitis. Bayesian analyses were done using the package *gemtc* [24] in the statistical software R [25]. Bayesian uninformative prior used was the default in the statistical analysis package (*gemtc*, R) [24], and Bayesian informative prior used was taken from a set of empirical priors [9], according to the comparison and outcome. Frequentist analyses were done using the package *netmeta* [26] in R.

parameter; this allows for the possibility of high and unrealistic levels of heterogeneity. As we have suggested, NMA protocols should include provisions for checking whether the assumptions made in their analyses hold, and if they might not, exploring results from alternative analyses that may prove more trustworthy.

9. The case of the single study

In our primary example of lidocaine versus placebo for out-of-hospital cardiac arrest, the direct comparison was based on a single study. Some may argue that if only one study informs the direct comparison, it is inappropriate to assume that the between-study heterogeneity is 0, and therefore, it should not be a problem that a network estimate for such a comparison has a wider CI than a direct estimate. In other words, we could learn about the between-study heterogeneity for comparisons for which there is only a single study from the entire network. Depending on the extent to which reviewers are willing to agree with this argument, they may deem trustworthy the results of NMAs in which a common heterogeneity was calculated using the whole network and estimates come from random effect models. We would argue that even when direct estimates come from a single study, if direct and indirect estimates are coherent (or when there is no indirect evidence), it is inappropriate to use a heterogeneity parameter that is poorly estimated in the network when it dramatically and implausibly widens the CI for the NMA estimate.

Parenthetically, others may argue that high-certainty direct evidence from a single study is not the same as high-certainty evidence from many studies. This consideration could provide some weight to the case for applying the heterogeneity parameter from the rest of the network to the single study. GRADE believes in replication—which is explicitly addressed when assessing inconsistency when rating the direct evidence—high-certainty direct evidence would have no limitations regarding the risk of bias, imprecision (which considers the optimal information size), applicability, and publication bias. A critical application of each of these considerations would make it unlikely that the addition of information from new studies would result in changes to the evidence that would lead to important differences in the consequent inferences. Thus, high-certainty evidence should be interpreted the same regardless of whether it comes from a single study or multiple studies, and reviewers who have obtained a high-certainty direct estimate should still deem it inappropriate to use a poorly estimated heterogeneity parameter in the context of an NMA.

10. Conclusions

Reviewers conducting an NMA of a sparse network should be aware of the potential issues that may arise from incorrect assumptions of the analysis and plan sensitivity analyses that lead to more trustworthy estimates of effects. Failing to do so will result in network estimates that have spuriously low certainty owing to limitations in the analysis methods rather than the evidence itself, making results less useful for patients, clinicians, and other decision-makers.

References

[1] Zarin W, Veroniki AA, Nincic V, Vafaei A, Reynen E, Motiwala SS, et al. Characteristics and knowledge synthesis approach for 456 network meta-analyses: a scoping review. BMC Med 2017;15:3.
[2] Efthimiou O, Mavridis D, Cipriani A, Leucht S, Bagos P, Salanti G. An approach for modelling multiple correlated outcomes in a network of interventions using odds ratios. Stat Med 2014;33:2275−87.
[3] Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health 2011;14:417−28.
[4] Mills EJ, Thorlund K, Ioannidis JP. Demystifying trial networks and network meta-analysis. BMJ 2013;346:f2914.
[5] 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.
[6] McLeod SL, Brignardello-Petersen R, Worster A, You J, Iansavichene A, Guyatt G, et al. Comparative effectiveness of antiarrhythmics for out-of-hospital cardiac arrest: a systematic review and network meta-analysis. Resuscitation 2017;121:90−7.
[7] Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. Stat Med 1996;15:2733−49.
[8] Lumley T. Network meta-analysis for indirect treatment comparisons. Stat Med 2002;21:2313−24.
[9] Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane database of systematic reviews. Int J Epidemiol 2012;41:818−27.
[10] Neupane B, Richer D, Bonner AJ, Kibret T, Beyene J. Network meta-analysis using R: a review of currently available automated packages. PLoS One 2014;9:e115065.
[11] 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 2018;93:36−44.
[12] 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.
[13] Thorlund K, Thabane L, Mills EJ. Modelling heterogeneity variances in multiple treatment comparison meta-analysis–are informative priors the better solution? BMC Med Res Methodol 2013;13:2.
[14] Petropoulou M, Nikolakopoulou A, Veroniki AA, Rios P, Vafaei A, Zarin W, et al. Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015. J Clin Epidemiol 2017;82:20−8.
[15] Song F, Xiong T, Parekh-Bhurke S, Loke YK, Sutton AJ, Eastwood AJ, et al. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. BMJ 2011;343:d4909.
[16] Deeks J, Higgins J. Chapter 9: analysing data and undertakink meta-analyses. In: JPT H, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration; 2011. Available from http://handbook.cochrane.org.

[17] Jansen JP, Crawford B, Bergman G, Stam W. Bayesian meta-analysis of multiple treatment comparisons: an introduction to mixed treatment comparisons. Value Health 2008;11:956−64.

[18] Hoaglin DC, Hawkins N, Jansen JP, Scott DA, Itzler R, Cappelleri JC, et al. Conducting indirect-treatment-comparison and network-meta-analysis studies: report of the ISPor task force on indirect treatment comparisons good research practices: part 2. Value Health 2011;14:429−37.

[19] Hong H, Carlin BP, Shamliyan TA, Wyman JF, Ramakrishnan R, Sainfort F, et al. Comparing Bayesian and frequentist approaches for multiple outcome mixed treatment comparisons. Med Decis Making 2013;33:702−14.

[20] Many reviews are systematic but some are more transparent and completely reported than others. PLoS Med 2007;4:e147.

[21] Booth A, Clarke M, Ghersi D, Moher D, Petticrew M, Stewart L. An international registry of systematic-review protocols. Lancet 2011;377:108−9.

[22] Silagy CA, Middleton P, Hopewell S. Publishing protocols of systematic reviews: comparing what was done to what was planned. JAMA 2002;287:2831−4.

[23] Kirkham JJ, Altman DG, Williamson PR. Bias due to changes in specified outcomes during the systematic review process. PLoS One 2010;5:e9810.

[24] van Valkenhoef G, Lu G, de Brock B, Hillege H, Ades AE, Welton NJ. Automating network meta-analysis. Res Synth Methods 2012;3:285−99.

[25] R Core Team. R. A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2016.

[26] Rücker G, Schwarzer G, Krahn U, König J. Netmeta: network meta-analysis using frequentist methods. R Package version 0.9-6, 2017: https://CRAN.R-project.org/package=netmeta.

[27] Urquhart O, Tampi MP, Pilcher L, Slayton R, Araujo MB, Espinoza L, et al. Nonrestorative treatments for caries: systematic review and network meta-analysis. J Dent Res 2018;5. 22034518800014.

[28] 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.