Check for updates

GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework

Romina Brignardello-Petersen,[1] Ivan D Florez,[1,2] Ariel Izcovich,[3] Nancy Santesso,[1] Glen Hazlewood,[4] Waleed Alhazanni,[1] Juan José Yepes-Nuñez,[5] George Tomlinson,[6,7] Holger J Schünemann,[1] Gordon H Guyatt,[1] on behalf of the GRADE working group

This article describes GRADE (grading of recommendations assessment, development and evaluation) guidance on how to draw conclusions from a network meta-analysis of interventions that includes individual randomised controlled trials addressing a single outcome. The guidance uses a minimally contextualised approach that avoids value judgments regarding the magnitude of intervention effects. The framework is based on two principles: interventions should be grouped in categories, from the most to the least effective or harmful; and the judgments that place interventions in such categories should simultaneously consider the estimates of effect, the certainty of the evidence, and the rankings. The framework includes five steps, which we describe and illustrate using an example. The framework is simple, methodologically sound, and flexible, allowing for modifications to resolve situations in which additional complexity or value judgments might be appropriate.

Optimal interpretation of results from a systematic review requires consideration of the magnitude of the estimates of effect, and the certainty of the evidence (confidence in evidence, quality of evidence).[1] In the context of network meta-analysis (NMA), users can consider measures of how interventions rank relative to one another. The larger the number of interventions, and thus the number of comparisons, the more complex and challenging the interpretation of NMA results becomes.

In networks looking at more than a few interventions, making inferences based on a simultaneous consideration of all possible comparisons is probably beyond the capacity of any individual and, as a result, summaries are necessary. In large networks, very seldom will an NMA, even with respect to one outcome, establish an intervention as clearly superior to all others.

Although there will always be an intervention that ranks higher than the others, focusing on ranking (generated through the surface under the cumulative ranking curve (in a bayesian analysis framework), or the P scores (in a frequentist analysis framework)) can be misleading for several reasons. Firstly, it tempts clinicians to focus on the apparent best intervention when that intervention might not be importantly different to others. Secondly, chance might easily explain apparent differences between ranks.[2] Thirdly, even if differences are real, they might be trivial in magnitude. Finally, and perhaps most importantly, rankings ignore certainty of evidence: a top ranked intervention might have only low or very low certainty evidence distinguishing it from comparators.[3 4] Thus, optimal interpretation of NMAs require alternatives beyond ranking.

The grading of recommendations assessment, development, and evaluation (GRADE) working group has presented guidance for evaluating the certainty of the evidence in NMAs,[5 6] how to avoid spurious judgments when addressing imprecision,[7] and how to assess incoherence.[8] The group, however, has so far not provided guidance on how to draw conclusions from NMAs.

This article describes how to draw conclusions from NMAs for a binary outcome or time-to-event outcome such as death, or a continuous outcome such as quality of life. Investigators can conduct this process using one of two approaches: a minimally contextualised framework and a partially contextualised framework. A minimally contextualised framework minimises value judgments regarding the magnitude of intervention effects. A partially contextualised approach will involve making such judgments.[9] This article focuses on the minimally contextualised approach.

SUMMARY POINTS

Network meta-analyses (NMA) rarely establish that one intervention is better than all others; reviewers should group interventions in categories, from the most to the least effective or the least to the most harmful

This article describes GRADE guidance on how to draw conclusions from NMA for one outcome using a transparent, straightforward, minimally contextualised approach that focuses on effect estimates and evidence certainty to classify interventions in groups from the most to the least effective or harmful

NMA GRADE users should use the new approach to ensure appropriate, informative conclusions that clinicians can easily understand

1

The presentation of the new approach in this article assumes familiarity with the basic concepts of NMA, implications of the four levels of certainty (high, moderate, low, and very low) of the GRADE evidence certainty system, and degrees of contextualisation. The named authors, under the auspices of the GRADE NMA project group, developed, tested, and refined the framework with feedback from the entire GRADE working group that ultimately approved the paper as GRADE guidance.

Methods

After a systematic survey of the literature showed that no methods have been proposed to draw conclusions from an NMA for one outcome that simultaneously consider the results from an NMA and the certainty in the evidence, a core team of experts (GHG, RB-P, JJY-N, IDF) in systematic review methodology and NMA developed an initial framework. The framework included guiding principles and a five step process, with aims to acknowledge the key features of complexity, as well as optimal simplicity and flexibility.

We obtained feedback regarding the initial framework from other experts in systematic review methodology, biostatisticians, and systematic reviewers, both with and without experience in NMA, and who were and were not members of the GRADE working group. As described below, participants provided feedback through small group sessions during GRADE working group meetings, semi-structured interviews with experts, and large group sessions during the GRADE working group meetings. Another core group of experts (GHG, RB-P, IDF, AI) considered the feedback and, after each round, made necessary changes to the framework.

At the GRADE working group meeting in Bogota, Colombia, in April 2018, participants formed small groups during two 1.5 hour sessions and provided the first round of feedback. About 30 members of the GRADE working group attended each session. During the sessions, we presented the initial framework through an example and opened the discussion to any feedback participants wished to provide.

For the second round, we contacted 10 experts and conducted 1 hour semi-structured interviews through an online video conference platform. In the interview, we presented the framework, revised after the first round, through an example and asked the experts to provide feedback regarding the guiding principles, five step process, and details from each step in the process.

For the third round, after modifications based on feedback from the second round, we presented the framework in a large group session at the GRADE working group meeting held in Manchester, UK, in September 2018. About 120 members of the GRADE working group attended this meeting. Again, we presented the framework through an example and opened the discussion to feedback.

After the three rounds of feedback, which resulted in minor modifications to the initial framework, we proceeded to use the framework in several NMAs to test feasibility and detect potential challenges. We used convenience sampling, applying the framework to NMAs from collaborating research groups and to NMAs that were then being developed by members of the GRADE NMA project group.[10-13] We tested our framework in examples with dichotomous and continuous outcomes, between six and 27 treatments, and diverse network geometry (including networks in which most of the comparisons were indirect and complex networks with several direct and indirect comparisons).

We made minor modifications after presenting the framework to the GRADE working group at a meeting held in Hamilton, Canada, in June 2019. We obtained approval to publish this framework as GRADE guidance at the GRADE working group meeting held in Adelaide, Australia, in November 2019.

Results

The minimally contextualised framework to draw conclusions from an NMA has two guiding principles and five steps. We describe and illustrate the simplest framework that we believe remains methodologically sound and which, based on our experience and testing with examples, is likely to work well in most instances. The framework also allows for flexibility and can be modified to accommodate additional complexity. Below, we describe the framework and some of the variations that reviewers might consider.

Guiding principles

Considering the insights outlined in the introduction, the framework to drawing conclusions from NMA is based on two principles:

- Interventions should be categorised (eg, those that are most effective, those with intermediate effectiveness, and those that are least effective). The number of categories will depend on the results of each NMA, and authors can modify the labels for each category according to the context.
- Judgments that place interventions in categories will rely primarily on the estimates of effect, and the certainty of the evidence supporting those estimates, and secondarily on the rankings. No single piece of information alone can determine whether a treatment is superior to others.

Use of the minimally contextualised framework to draw conclusions from network meta-analyses

The process of drawing conclusions has five steps. Before implementing these steps, reviewers must rate the certainty of the evidence of each network estimate.[5][6] Below, we describe and illustrate each step using an NMA of pharmacological and nutritional interventions for treating acute diarrhoea and gastro-enteritis in children.[14] We present two other examples in the appendix.[12][15]

The primary outcome of the paediatric gastroenteritis review[14] was diarrhoea duration, measured in hours. Because the interventions are expected to decrease

diarrhoea duration, our discussion focuses on this beneficial outcome. The framework, however, can also be applied to harm or safety outcomes. The 138 eligible studies in the paediatric gastroenteritis review recruited 20 256 children and assessed 27 interventions. The network has a complex geometry (fig 1) with 62 direct comparisons; the remaining 289 comparisons have only indirect evidence.

Step 1: Choose reference intervention and decision threshold

The process begins with choosing a reference intervention, which should be the most connected to other interventions in the network. Network estimates that are calculated with direct evidence are more likely to provide higher certainty evidence and more precise estimates of effect than those based only on indirect evidence. Because of the higher likelihood of differentiating between treatments when there is higher or moderate certainty evidence rather than low or very low certainty evidence, anchoring the process using evidence with higher certainty is most appropriate when drawing conclusions from NMAs. In other words, the categorisation is more likely to be informative if it is anchored to higher certainty evidence.

In the NMA of interventions for acute diarrhoea in children,[14] the reference intervention was standard treatment that included arms characterised as "no active treatment," "placebo," or as "only oral rehydration solution." To claim that one treatment is better than another, review authors must choose a decision threshold. To keep a framework minimally contextualised, reviewers could choose no effect

(eg, a relative effect of 1.0, and absolute difference of 0) as the threshold—that is, one treatment in a comparison will be considered superior only if the 95% confidence or credible interval excludes the null. In more contextualised approaches, this value might be a minimally important difference or an importance threshold. Choosing a relative effect of 1.0 maximises the possibility of differentiating among treatments, but could lead to claiming that one intervention is better than another when the magnitude of the difference is not important to patients. In our example, we chose a change in diarrhoea duration of 3 hours as the threshold to decide whether the effect of one intervention differs from another.

Step 2: First classification of interventions based on comparison with reference

In this step, reviewers use the 95% confidence or credible interval of the estimate of effect comparing each of the interventions against the reference. If this interval crosses the decision threshold, the intervention will remain in the same group as the reference. If, on the other hand, the interval does not cross the decision threshold, the intervention can be classified as more or less effective than the reference, depending on which side of the threshold the interval lies (fig 2).

This classification is likely to result in having two groups—interventions not convincingly different from the reference and those more or less effective than the reference. However, the process could result in three groups: interventions not different from the reference, those more effective than the reference, and those less effective than the reference. For harm outcomes, interventions can be classified as those not different from the reference, those less harmful than the reference, and those more harmful than the reference.

In the NMA of interventions for acute diarrhoea in children,[14] for the outcome of diarrhoea duration, we found interventions not convincingly different from placebo and those more effective than placebo (box 1). To facilitate description of the process, we will label these groups as category 0, and category 1, respectively. When the threshold is an important difference (rather than the null effect), this step requires absolute values that (as in this example) will be the natural report for continuous outcomes. The same, however, is not true for binary outcomes in which the NMA will yield estimates of relative effect that then need translation into absolute effect.

Translation to absolute effects is necessary because judgments of importance cannot be made on the basis of relative effects. For example, a 50% relative reduction with a baseline risk of 2% represents a 1% absolute risk reduction that might be considered as unimportant. That same 50% relative risk reduction, in the setting of a baseline risk of 40%, represents a 20% absolute risk reduction that might be judged as very important.
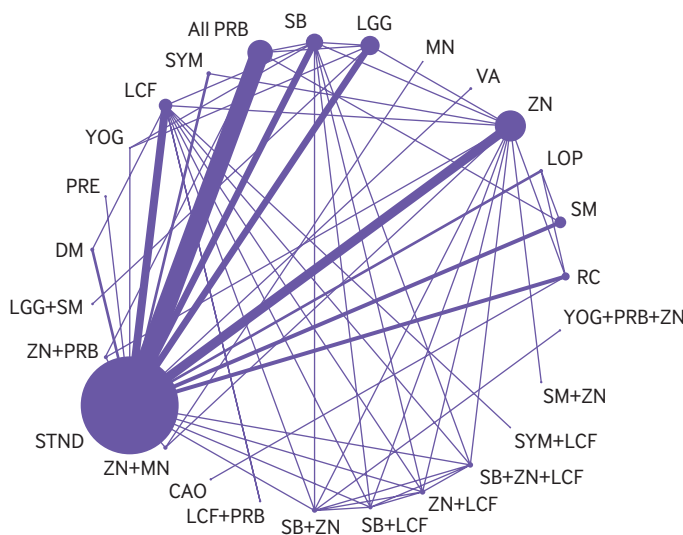

Fig 1 | Network plot of interventions for treating acute diarrhoea and gastroenteritis in children, for the outcome diarrhoea duration, based on a paediatric gastroenteritis review.[14] PRB=all probiotics; SB=*Saccharomyces boulardii*; LGG=*Lactobacillous rhamnosus GG*; MN=micronutrients; VA=vitamin A; ZN=zinc; LOP=loperamide; SM=smectite; RC=racecadotril; YOG=yoghurt; SYM=symbiotics; LCF=lactose free formula; CAO=Kaolin-pectin; STND=placebo or standard care; DM=diluted milk; PRE=prebiotics
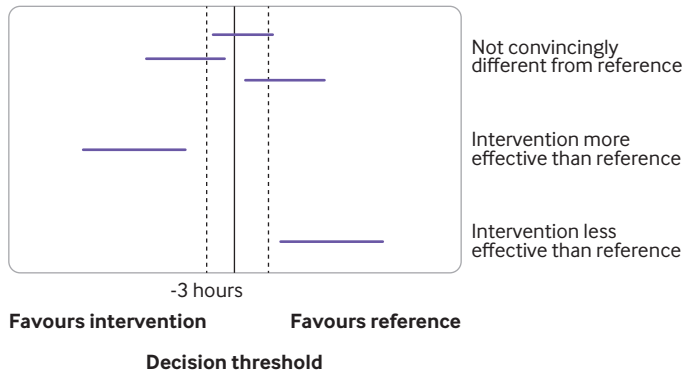
Fig 2 | Classification of interventions in network meta-analyses, based on comparison with the reference intervention

Step 3: Second classification based on comparisons between pairs of interventions

In this step, NMA authors compare the interventions classified as more effective than the reference against each other by examining whether the confidence or credible interval of their estimate of effect crosses the decision threshold. The decision threshold is the same one used for the first classification. If any intervention proves more effective than another category 1

intervention, that intervention is moved to a higher rated group (that is, category 2).

In the gastroenteritis review,[14] because the mean difference of the comparison between zinc + micronutrients versus zinc is 0.63 hours (95% confidence interval −13.20 to 14.56; mean difference of ≤3 hours remains plausible), zinc + micronutrients remain in the same category as zinc (that is, category 1). Because the comparison between *Saccharomyces boulardii* + zinc versus zinc has a mean difference of −21.55 (−33.66 to −9.38; mean difference of ≤3 hours is implausible), reviewers will classify *S boulardii* + zinc in category 2.

Subsequently, reviewers can implement this same step to differentiate among interventions in category 2 (if there is an intervention in category 2 superior to at least one other, it would move to category 3) until no new groupings can be made. So far, we have not encountered an instance in which a category 3 intervention exists. In the NMA of interventions for acute diarrhoea in children,[14] interventions in category 1 could be separated into two groups, thus creating a category 2; reviewers could not make any further differentiation (box 2).

Step 4: Separate interventions into two main groups according to certainty of evidence

In this step, reviewers identify the certainty of the evidence for each of the interventions when compared against the reference, and categorise interventions as those with high or moderate certainty evidence when compared with the reference, and those with low or very low certainty evidence when compared with the reference. Table 1 shows how reviewers classified the 27 interventions for acute diarrhoea in children after this step, in our NMA example.[14]

Step 5: Checking consistency with pairwise comparisons and rankings

In this final step, reviewers examine the pairwise comparisons not previously considered to make sure that the classification is not inconsistent with these other comparisons. They can also look at the ranking to ensure that those interventions ranked highest were among the most effective.

Compelling evidence of limitations of the classification before step 5 can result in modifications to the classification. For example, an estimate with high or moderate certainty showing that an intervention placed in category 0 is more effective than one placed in category 1 would probably lead to that intervention moving to category 2. No such examples were found in the NMA of interventions for acute diarrhoea in children[14] (table 1); indeed, the nature of the statistical NMA process makes such a situation very unlikely. The classification proved otherwise consistent with the rankings. For example, SUCRA (surface under the cumulative ranking curve) values tended to be highest for category 2 interventions (among the most effective), lower for category 1 interventions, and lowest for category 0 interventions (among the least effective).

Box 1: First classification of interventions based on comparison with reference, for the outcome of diarrhoea duration (in a network meta-analysis of interventions for acute diarrhoea in children[14])

Not convincingly different than placebo (category 0)
- Prebiotics
- *Saccharomyces boulardii* + zinc + lactose-free formula
- Yoghurt + probiotics + zinc
- Lactose-free formula + probiotics
- *S boulardii* + lactose-free formula
- Vitamin A
- Kaolin-pectin
- Micronutrients
- Standard treatment or placebo
- Diluted milk
- Yoghurt

More effective than placebo (category 1)*
- *S boulardii* + zinc
- Smectite + zinc
- *Lactobacillous rhamnosus GG* + smectite
- Zinc + probiotics
- Symbiotics
- Zinc + lactose-free formula
- Zinc
- Loperamide
- Zinc + micronutrients
- Symbiotics + lactose-free formula
- Smectite
- *L rhamnosus GG*
- Probiotics
- Racecadotril
- *S boulardii*
- Lactose-free formula

Box 2: Second classification based on pairwise comparisons, for the outcome of diarrhoea duration (in network meta-analysis of interventions for acute diarrhoea in children[14])

Not convincingly different than placebo (category 0)
- Prebiotics
- *Saccharomyces boulardii* + zinc + lactose-free formula
- Yoghurt + probiotics + zinc
- Lactose-free formula + probiotics
- *S boulardii* + lactose-free formula
- Vitamin A
- Kaolin-pectin
- Micronutrients
- Standard treatment or placebo
- Diluted milk
- Yoghurt

Intermediate—more effective than placebo (category 1)
- Symbiotics
- Zinc + lactose-free formula
- Zinc
- Loperamide
- Zinc + micronutrients
- Symbiotics + lactose-free formula
- Smectite
- *Lactobacillous rhamnosus GG*
- Probiotics
- Racecadotril
- *S boulardii*
- Lactose-free formula

More effective than at least one category 1 intervention (category 2)
- *S boulardii* + zinc
- Smectite + zinc
- *L rhamnosus GG* + smectite
- Zinc + probiotics

Although this step is unlikely to change the final categorisation if the assumptions of NMA are met, reviewers should consider the information provided by all pairwise comparisons that do not involve the reference. Steps 1-4 of our framework do not consider these comparisons, and a safeguard against any possible mistake is advisable. Appendix 1 shows an example in which step 5 could result in a modification of the categorisation based on the pairwise comparisons and rankings.

After these five steps, review authors can communicate their findings using language that GRADE suggests to convey higher certainty (interventions are among the most effective) and language appropriate for lower certainty (interventions may be among the most effective).[16] In our example NMA of interventions for acute diarrhoea in children,[14] the conclusions are as follows:

- *S Boulardii* + zinc and smectite + zinc are among the most effective interventions to reduce diarrhoea
- Symbiotics, zinc + lactose-free formula, zinc, loperamide, zinc + micronutrients are among the interventions with intermediate effectiveness to reduce diarrhoea

- Prebiotics are among the least effective interventions to reduce diarrhoea
- *Lactobacillous rhamnosus GG* + smectite and zinc + probiotics could be among the most effective interventions to reduce diarrhoea
- Symbiotics + lactose-free formula, smectite, *L rhamnosus GG*, all probiotics, racecadotril, *S boulardii*, yoghurt, and lactose-free formula could be among the interventions with intermediate effectiveness to reduce diarrhoea
- All the other interventions could be among the least effective interventions to reduce diarrhoea

Possible modifications to the process

This minimally contextualised framework allows reviewers to classify interventions while making few value judgments. Reviewers might, however, find some criteria questionable, but the framework can easily accommodate modifications. For example:

- We have pointed out that the decision threshold might be no difference or a threshold of minimal importance. In this example, we chose a threshold of at least 3 hours of reduction in the duration of diarrhoea. Different thresholds can be used; for example, reviewers might judge that a reduction in the number of hours of diarrhoea duration less than 3 hours is trivial and therefore use a threshold of 6 hours. If this threshold was used, lactose-free formula (mean difference −12.50 hours (95% confidence interval −19.04 to −5.99)), zinc + macronutrients (−17.76 hours (−31.77 to −4.13)), and loperamide (−17.79 hours (−30.35 to −5.65)) would have been classified as no different than placebo. Reviewers using a decision threshold different than the null value should choose such thresholds using absolute estimates of effect that, for binary outcomes, requires transforming from relative to absolute effects (a relative effect of 0.8 could be a large or small effect depending on the baseline risk).
- Reviewers can modify rules to determine whether interventions other than the reference are superior to another (step 3 in our five step process). Reviewers can, for instance, require moderate or high certainty evidence of a difference to move interventions from category 1 to category 2.
- Reviewers might want to differentiate among the levels of certainty of the evidence, and not group high and moderate interventions and low and very low interventions into the same categories.

Discussion

We have described a framework to draw conclusions from an NMA in which reviewers can classify interventions in categories: those most effective to those least effective for one outcome, each supported by higher versus lower certainty evidence. We describe this framework as minimally contextualised because it requires few value judgments regarding the magnitude of effects and the trade-off between desirable and

Table 1 | Final classification of 27 interventions, based on network meta-analysis of interventions for acute diarrhoea in children[14]

Certainty of the evidence, and classification* of intervention	Intervention†	Intervention v standard treatment or placebo (mean difference (95% credible interval))	Surface under the cumulative ranking curve
High certainty (moderate to high certainty evidence)			
Category 2: among the most effective	*Saccharomyces boulardii* + zinc (M)	−39.45 (−52.5 to −26.7)	0.92
	Smectite + zinc (M)	−35.63 (−57.6 to −13.2)	0.88
Category 1: inferior to the most effective, or superior to the least effective	Symbiotics (H)	−26.26 (−36.1 to −16.2)	0.77
	Zinc + lactose-free formula (M)	−21.37 (−36.5 to −6.1)	0.61
	Zinc (M)	−18.38 (−23.4 to −13.5)	0.50
	Loperamide (M)	−17.79; (−30.4 to −5.7)	0.46
	Zinc + micronutrients (M)	−17.76 (−31.8 to −4.1)	0.46
Category 0: among the least effective	Prebiotics (M)	−15.32 (−42.8 to 12.0)	0.38
Low certainty (low to very low certainty evidence)			
Category 2: might be among the most effective	*Lactobacillous rhamnosus GG* + smectite (VL)	−51.08 (−64.3; to −37.9)	1.00
	Zinc + probiotics (L)	−29.39 (−40.3 to −18.6)	0.81
Category 1: might be inferior to the most effective or superior than the least effective	Symbiotics + lactose-free formula (VL)	−32.11 (−53.0 to −11.3)	0.85
	Smectite (VL)	−23.90 (−30.8 to −17.0)	0.69
	L rhamnosus GG (L)	−22.74 (−28.8 to −16.7)	0.65
	Probiotics (L)	−19.36 (−23.7 to −15.1)	0.54
	Racecadotril (L)	−17.19 (−24.7 to −9.8)	0.46
	S boulardii (L)	−16.48 (−23.3 to −9.7)	0.42
	Lactose-free formula (VL)	−12.50 (−19.0 to −6.0)	0.31
Category 0: might be among the least effective	*S boulardii* + zinc + lactose-free formula (L)	−16.74 (−36.1 to 2.7)	0.42
	Yoghurt + probiotics + zinc (VL)	−15.63 (−56.8 to 26.6)	0.38
	Lactose-free formula + probiotics (VL)	−13.27 (−36.0 to 9.2)	0.31
	S boulardii + lactose-free formula (VL)	−12.32 (−30.0 to 6.0)	0.27
	Vitamin A (VL)	−5.95 (−21.4 to 9.3)	0.19
	Kaolin-pectin (VL)	−5.32 (−33.8 to 22.8)	0.15
	Micronutrients (L)	−0.68 (−33.3 to 32.8)	0.08
	Standard treatment or placebo	—	0.08
	Yoghurt (VL)	−16.43 (−30.5 to −2.1)	0.42
	Diluted milk (VL)	3.02 (−14.3 to 8.4)	0.04

*Categories do not inform value judgments about the importance of the effects. A suggested format of presentation could include different colours and shades.
†Letters in brackets represent the certainty of evidence for each intervention when compared with the reference: H=high certainty evidence; M=moderate; L=low; VL=very low.

undesirable effects of the interventions. The number of resulting categories depends on the evidence available, and it will be influenced by how many interventions are included in the NMA, how they compare to one another, and the decision threshold. Reviewers should apply this framework after they have assessed the certainty of the evidence for each network estimate using the GRADE approach for NMA.[6]

Adopting a minimally contextualised approach allows easy application of the framework across a variety of contexts and facilitates the process of drawing conclusions from NMA. Including the possibility of a reference intervention most connected in the network, rather than a less connected placebo or no treatment, facilitates a focus on the highest certainty evidence when drawing conclusions. Basing the differentiation of interventions according to certainty of evidence into two categories (high and moderate certainty v low and very low certainty) puts a premium on trustworthy evidence; focusing that differentiation on comparisons with the reference further simplifies the process.

Our framework can apply to different bodies of evidence where certainty is assessed by traditional GRADE principles. Our examples are NMAs in which eligible studies had a randomised or quasi-randomised design. Applying our framework to bodies of evidence composed of observational studies would also be possible, although the number of resulting categories is likely to be smaller, given how unlikely it is that these bodies of evidence are judged at high or moderate certainty.

Reviewers might correctly point out a limitation of our approach, which is the focus on the reference to the relative exclusion of non-reference paired comparisons. That neglect is not, however, complete. The third step of this process requires reviewers to focus on pairwise comparisons that do not involve the reference. Moreover, the last step demands that reviewers verify that the final categorisation is not inconsistent with the comparisons between interventions other than the reference.

Another potential limitation of our approach is the lack of adjustment for multiplicity. How seriously multiplicity might compromise inferences remains uncertain, and no satisfactory or widely accepted strategy exists for dealing with the problem. Therefore, we did not consider adjusting for multiplicity in our framework. In addition, the simplicity of our framework, which is part of a complex systematic review process, was one of our priorities.

Some reviewers might believe that drawing these conclusions without the necessary contextualisation is not appropriate. Reviewers have, however, successfully applied this minimally contextualised framework in several published NMAs[14 17-21] or in process of publication.[22] In all these examples, clinical experts closely involved in the systematic review and in the peer review process have been satisfied with the output of the framework and found it helpful.

Table 2 | Similarities and differences between the GRADE minimally and partially contextualised frameworks for drawing conclusions from NMA

	Minimally contextualised framework	Partially contextualised framework
Similarities		
Overall objective	To group interventions in different categories according to their effects	
Information considered	Estimates of effect, certainty in the evidence, rankings	
General inferences	Interventions placed in higher categories are likely to be more effective than interventions placed in lower categories	
Differences		
Primary statistic used to categorise interventions	Confidence interval	Point estimate
Value placed on imprecision	Higher than other limitations of the evidence	Same as other limitations of the evidence
Resulting categories	From the most to the least effective/ harmful	From large to trivial benefit or harm

When concerns regarding contextualisation remain salient, use of a partially contextualised framework, described elsewhere, could be helpful.[23] Table 2 summarises the similarities and differences between the two approaches. The two approaches share overall objectives, information considered, and general inferences. The minimally contextualised approach we have described in this paper relies on the position of the confidence interval with regard to a threshold to categorise interventions, thus emphasising issues of precision, and categorises interventions from the most to least beneficial or harmful (table 1). The partially contextualised approach described elsewhere[23] focuses on point estimates to categorise interventions and classifies interventions according to the magnitude of benefit and harm from large to trivial.

The differences between the two frameworks could cause two interventions to be categorised the same using one framework and differently using the other framework. For example, in the NMA of interventions for acute diarrhoea,[14] *S boulardii* + zinc was classified among the most effective, and symbiotics was classified as inferior to the most effective. This classification was because no convincing evidence indicated that symbiotics was better than any other intervention in its category as determined by the association between the confidence intervals and the decision threshold. When using a partially contextualised framework, these two interventions are in the same category because of the association between each of their point estimates and the thresholds of magnitude of effect.[23]

Conclusion

This minimally contextualised framework facilitates the development of conclusions from NMA, considering all the crucial information. The framework places a high emphasis on simplicity and applicability across different contexts, while minimising the need of judgments that might be context specific. In addition, its flexibility allows reviewers to modify the framework as appropriate to ensure that they reach sensible conclusions.

Author affiliations
[1]Department of Health Research Methods, Evidence and Impact, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4L8, Canada

[2]Department of Pediatrics, School of Medicine, University of Antioquia, Medellín, Colombia

[3]Internal Medicine Service, German Hospital, Buenos Aires, Argentina

[4]Department of Medicine, Cumming School of Medicine, University of Calgary, Calgary, AB, Canada

[5]School of Medicine, Universidad de los Andes, Bogota, Colombia

[6]Institute of Health Policy, Management, and Evaluation, University of Toronto, Toronto, ON, Canada

[7]Biostatistics Research Unit, University Health Network, Toronto, ON, Canada

1   Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation-determinants of a recommendation's direction and strength. *J Clin Epidemiol* 2013;66:726-35. doi:10.1016/j.jclinepi.2013.02.003

2   Trinquart L, Attiche N, Bafeta A, Porcher R, Ravaud P. Uncertainty in treatment rankings: reanalysis of network meta-analyses of randomized trials. *Ann Intern Med* 2016;164:666-73. doi:10.7326/M15-2521

3   Brignardello-Petersen R, Guyatt GH. β-Blockers in heart failure--are all created equal? *Pol Arch Med Wewn* 2013;123:204-5. doi:10.20452/pamw.1720

4   Neumann I, Brignardello-Petersen R, Guyatt G. ACP journal club. Review: Novel oral anticoagulants reduce stroke more than ASA in nonvalvular atrial fibrillation. *Ann Intern Med* 2014;160:JC3. doi:10.7326/0003-4819-160-4-201402180-02003

5   Puhan MA, Schünemann HJ, Murad MH, et al, GRADE Working Group. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ* 2014;349:g5630. doi:10.1136/bmj.g5630

6   Brignardello-Petersen R, Bonner A, Alexander PE, et al, GRADE Working Group. Advances in the GRADE approach to rate the

certainty in estimates from a network meta-analysis [correction in: *J Clin Epidemiol* 2018;98:162]. *J Clin Epidemiol* 2018;93:36-44. doi:10.1016/j.jclinepi.2017.10.005

7   Brignardello-Petersen R, Murad MH, Walter SD, et al, GRADE Working Group. GRADE approach to rate the certainty from a network meta-analysis: avoiding spurious judgments of imprecision in sparse networks. *J Clin Epidemiol* 2019;105:60-7. doi:10.1016/j.jclinepi.2018.08.022

8   Brignardello-Petersen R, Mustafa RA, Siemieniuk RAC, et al, GRADE Working Group. GRADE approach to rate the certainty from a network meta-analysis: addressing incoherence. *J Clin Epidemiol* 2019;108:77-85. doi:10.1016/j.jclinepi.2018.11.025

9   Hultcrantz M, Rind D, Akl EA, et al. The GRADE working group clarifies the construct of certainty of evidence. *J Clin Epidemiol* 2017;87:4-13. doi:10.1016/j.jclinepi.2017.05.006

10 Rochwerg B, Alhazzani W, Sindi A, et al, Fluids in Sepsis and Septic Shock Group. Fluid resuscitation in sepsis: a systematic review and network meta-analysis. *Ann Intern Med* 2014;161:347-55. doi:10.7326/M14-0178

11 Rochwerg B, Neupane B, Zhang Y, et al. Treatment of idiopathic pulmonary fibrosis: a network meta-analysis. *BMC Med* 2016;14:18. doi:10.1186/s12916-016-0558-x

12 Alhazzani W, Alshamsi F, Belley-Cote E, et al. Efficacy and safety of stress ulcer prophylaxis in critically ill patients: a network meta-analysis of randomized trials [correction in: *Intensive Care Med* 2018;44:277-278]. *Intensive Care Med* 2018;44:1-11. doi:10.1007/s00134-017-5005-8

13 Hazlewood GS, Barnabe C, Tomlinson G, Marshall D, Devoe DJ, Bombardier C. Methotrexate monotherapy and methotrexate combination therapy with traditional and biologic disease modifying anti-rheumatic drugs for rheumatoid arthritis: A network meta-analysis. *Cochrane Database Syst Rev* 2016;(8):CD010227. doi:10.1002/14651858.CD010227.pub2

14 Florez ID, Veroniki AA, Al Khalifah R, et al. Comparative effectiveness and safety of interventions for acute diarrhea and gastroenteritis in children: A systematic review and network meta-analysis. *PLoS One* 2018;13:e0207701. doi:10.1371/journal.pone.0207701

15 Hazlewood GS, Barnabe C, Tomlinson G, Marshall D, Devoe D, Bombardier C. Methotrexate monotherapy and methotrexate combination therapy with traditional and biologic disease modifying

antirheumatic drugs for rheumatoid arthritis: abridged Cochrane systematic review and network meta-analysis. *BMJ* 2016;353:i1777. doi:10.1136/bmj.i1777

16 Santesso N, Glenton C, Dahm P, et al, GRADE Working Group. GRADE guidelines 26: informative statements to communicate the findings of systematic reviews of interventions. *J Clin Epidemiol* 2020;119:126-35. doi:10.1016/j.jclinepi.2019.10.014

17 Niño-Serna LF, Acosta-Reyes J, Veroniki A-A, Florez ID. Antiemetics in children with acute gastroenteritis: a meta-analysis. *Pediatrics* 2020;145:e20193260. doi:10.1542/peds.2019-3260

18 Noori A, Busse JW, Sadeghirad B, et al. Individual opioids, and long- versus short-acting opioids, for chronic noncancer pain: Protocol for a network meta-analysis of randomized controlled trials. *Medicine (Baltimore)* 2019;98:e17647. doi:10.1097/MD.0000000000017647

19 Zeng L, Qasim A, Neogi T, et al. Comparative efficacy and safety of pharmacological interventions in patients experiencing a gout flare: a systematic review and network meta-analysis. *Arthritis Care Res* 2020. doi:10.1002/acr.24402

20 Lytvyn L, Zeraatkar D, Anbari AB, et al. Conservative intervention strategies for adult cancer-related lymphedema: a systematic review and network meta-analysis. *Oncol Nurs Forum* 2020;47:E171-89. doi:10.1188/20.ONF.E171-E189

21 Morgan R, Preidis G, Kashyap P, Weizman A, Sadeghirad B. Probiotics for prevention of mortality and morbidity in preterm infants: a systematic review and network meta-analysis of randomized trials. *Gastroenterology* 2020;159:467-80. doi:10.1053/j.gastro.2020.05.096

22 Hutton B, Hersi M, Cheng W, et al. Comparison of the effects of natural health products, pharmacologics, physical activity, psychologic therapy and combination interventions on hot flash frequency and intensity in patients with a history of breast and prostate cancer: a systematic review with meta-analyses. *J Oncol Nurs* [forthcoming].

23 Brignardello-Petersen R, Izcovich A, Rochwerg B, et al, on behalf of the GRADE working group. GRADE approach to drawing conclusions from a network meta-analysis using a partially contextualized framework. *BMJ* 2020;371:m3907.

Web appendix: Supplementary materials