A guide and pragmatic considerations for applying GRADE to network meta-analysis

Ariel Izcovich,1 Derek K Chu,2,3 Reem A Mustafa,2,4 Gordon Guyatt,2,3 Romina Brignardello-Petersen2

Assessing the certainty of evidence from network meta-analyses using the GRADE (grading of recommendations, assessment, development, and evaluations) approach requires not only a thorough understanding of the methods but also substantial workload for raters. This article describes how implementing practical strategies (including rule setting and automation) can facilitate efficient application of the GRADE approach to rating certainty of evidence in network meta-analyses while maintaining rigor. This article describes a stepwise strategy for proceeding through the process, including assessment of the certainty of direct, indirect, and NMA estimates; developing directions for each step in the process; implementation of rules to improve robustness of information appraisal to reduce workload; and alternatives for automation of some of the required steps. The presented approach includes a detailed description of every step of the process supported by figures, tables, and real life examples. To facilitate implementation of this process, a spreadsheet that incorporates automation of several of the steps described is also provided (https://www.covid19lnma.com/).

Network meta-analysis (NMA) allows assessing the comparative effectiveness of multiple interventions by combining direct and indirect evidence in one statistical model, resulting in estimates of effect comparing every pair of interventions included in the network—even if they have not been compared directly in trials.1 Assessing the certainty of the evidence (also known as quality of the evidence, and confidence in the estimates of effects) from NMAs is crucial for interpreting those estimates and moving from evidence to decision. The GRADE (grading of recommendations, assessment, development, and evaluations) working group has provided guidance for assessing the certainty of the evidence and drawing conclusions from NMAs.2-8 The approach considers the certainty of all direct, indirect, and network (also known as mixed) estimates between interventions (nodes) included in the network (fig 1). Implementing the GRADE approach for NMA requires an understanding of the methods, and an awareness that the larger the network, the much larger is the increase in workload. For example, in a network with four interventions, there are six comparisons for which to assess the certainty of evidence, whereas a network with 10 interventions has 45 comparisons. To complete the GRADE certainty of the evidence assessment, raters need to make between 12 and 13 judgments for every comparison (fig 1). This workload represents a challenge in implementing GRADE in systematic reviews that use NMA.

In this article, we present pragmatic strategies for applying GRADE for NMA, which we developed through our experience conducting a living systematic review and NMA of randomised trials looking at the prophylaxis and treatment of covid-19.9-11 As of December 2021, this large NMA included 57 nodes and 1597 comparisons for the outcome mortality in the drug treatments NMA. We developed this process to seek the most efficient way of performing all the necessary steps to assess the certainty of evidence following the GRADE approach by, among other strategies, taking advantage of previously performed steps (eg, once the certainty of the evidence for a direct comparison is determined, there is no need to reassess the same comparison when considered as indirect evidence) and identifying optional steps (eg, assessing imprecision for direct and indirect estimates is not always necessary). We then developed a spreadsheet

<table>
<thead>
<tr>
<th>SUMMARY POINTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Assessing the certainty of the evidence from network meta-analyses (NMAs) using the GRADE approach requires a thorough understanding of the methods and a substantial workload.</td>
</tr>
<tr>
<td>This article describes a step-by-step process for assessing the certainty of the evidence of each comparison in an NMA.</td>
</tr>
<tr>
<td>Practical strategies are also described, such as rule setting and automation, which can facilitate implementation of the GRADE approach in the entire network.</td>
</tr>
<tr>
<td>By following the proposed approach and incorporating tools that facilitate the tasks, reviewers will achieve a substantial reduction in the workload without compromising the rigor.</td>
</tr>
</tbody>
</table>
Assessing the certainty of evidence for one comparison in a network meta-analysis

This diagram explains the process of assessing a single comparison within a network meta-analysis. For a summary of the process for assessing certainty in a whole network, please see supplementary table 1.

**Table 1 | Timeline and focus of publications describing GRADE guidance for assessing the certainty of evidence from network meta-analysis**

<table>
<thead>
<tr>
<th>First author and year</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Puhan 2014*</td>
<td>Describes the principles and process for assessing the certainty of evidence from network meta-analysis using the GRADE approach.</td>
</tr>
<tr>
<td>Brignardello-Petersen 2018*</td>
<td>Describes conceptual advances of the GRADE approach for assessing the certainty of evidence from network meta-analysis. Focuses on strategies to improve efficiency of the process and on incoherence.</td>
</tr>
<tr>
<td>Brignardello-Petersen 2019*</td>
<td>Provides guidance for avoiding spurious judgments on imprecision due to counterintuitive results obtained in sparse networks.</td>
</tr>
<tr>
<td>Brignardello-Petersen 2019*</td>
<td>Provides detailed guidance for assessing incoherence.</td>
</tr>
<tr>
<td>Brignardello-Petersen 2020*</td>
<td>Presents and describes a strategy for drawing conclusions from network meta-analysis using a partially contextualised framework.</td>
</tr>
<tr>
<td>Brignardello-Petersen 2021*</td>
<td>Provides detailed guidance for assessing imprecision of network meta-analysis estimates.</td>
</tr>
</tbody>
</table>

GRADE = grading of recommendations, assessment, development, and evaluations.

Fig 1 | Assessing the certainty of evidence for one comparison in a network meta-analysis. RoB = risk of bias; Inc = inconsistency; Ind = indirectness; Pub = publication bias
three presented examples (examples 1-3) and one additional example (example 4).

**Assessing the certainty of the direct estimate**

Reviewers perform the assessment of direct estimates to inform the assessment of the network estimate in the same way as they would conduct the GRADE approach for traditional meta-analysis with one exception, the assessment of imprecision of the direct estimate (steps A1 and A2 in fig 1).8

**Risk of bias, inconsistency, indirectness, and publication bias**

*Step A1*

Reviewers must assess risk of bias, inconsistency, indirectness, and publication bias for all comparisons for which direct estimates are available. For example, for one of our outcomes of interest (mortality), from 1597 possible comparisons, 82 had a direct estimate that required assessment. Reviewers can follow extensive published guidance on assessment of the domains.16-20 When completed, a rating to inform the NMA estimate (direct preliminary certainty rating omitting imprecision assessment) becomes available (step A1 in fig 1).

**Final rating of the direct estimate**

*Step A2*

Reviewers need to assess imprecision of the direct estimate to obtain a final rating of the direct estimate (step A2 in fig 1).21-23 Obtaining a final rating for direct estimate is not always necessary (as will become evident at step D). Reviewers can choose to skip—or ultimately delay—step A2 and complete it later, if deemed necessary. Figure 2 illustrates the process for the assessment of the certainty of evidence for the direct estimate, including this step, for three comparisons on three different outcomes of interest.

**Practical considerations**

When dealing with large networks, reviewers can create spreadsheets to automate some steps of the process. For example, in the supplementary spreadsheet provided (supplementary file 1), preliminary certainty ratings to inform NMA are automatically calculated based on reviewers’ ratings of the four assessed domains (risk of bias, inconsistency, indirectness, and publication bias).

The ratings of certainty of the evidence of direct comparisons are the building blocks for both the assessment of indirect estimates and the network estimates. In addition, each direct comparison rating can inform the certainty rating of many indirect and network comparisons. Therefore, reviewers will benefit by starting the process with rating the certainty of all direct estimates.

### Table: Final certainty rating of direct estimates

<table>
<thead>
<tr>
<th>Comparison</th>
<th>Risk-of-bias assessment</th>
<th>Inconsistency assessment</th>
<th>Indirectness assessment</th>
<th>Publication bias assessment</th>
<th>Direct certainty rating to inform the network estimate*</th>
<th>Imprecision assessment*</th>
<th>Final certainty rating of direct estimate*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Direct estimate on systemic corticosteroids v standard of care for covid-19. Outcome: Mortality</td>
<td>Serious methodological limitations</td>
<td>No serious inconsistency</td>
<td>No serious indirectness</td>
<td>Publication bias undetected</td>
<td>Moderate certainty (++)</td>
<td>Serious imprecision (++)</td>
<td>Low certainty (++)</td>
</tr>
<tr>
<td>Rate down one level</td>
<td>Do not rate down</td>
<td>Do not rate down</td>
<td>Do not rate down</td>
<td>Moderate certainty</td>
<td>Serious imprecision</td>
<td>Low certainty</td>
<td></td>
</tr>
<tr>
<td>Direct estimate on azithromycin v standard of care for covid-19. Outcome: Time to symptom resolution</td>
<td>Serious methodological limitations</td>
<td>Serious inconsistency</td>
<td>No serious indirectness</td>
<td>Publication bias undetected</td>
<td>Low certainty (++)</td>
<td>No serious imprecision (++)</td>
<td>Low certainty (++)</td>
</tr>
<tr>
<td>Rate down one level</td>
<td>Rate down one level</td>
<td>Do not rate down</td>
<td>Do not rate down</td>
<td>Low certainty (++)</td>
<td>No serious imprecision</td>
<td>Low certainty (++)</td>
<td></td>
</tr>
<tr>
<td>Direct estimate on ivermectin v hydroxychloroquine for covid-19. Outcome: Mechanical ventilation</td>
<td>No serious methodological limitations</td>
<td>No serious inconsistency</td>
<td>No serious indirectness</td>
<td>Publication bias undetected</td>
<td>High certainty (++++)</td>
<td>Very serious imprecision (++++)</td>
<td>Low certainty (++)</td>
</tr>
<tr>
<td>Do not rate down</td>
<td>Do not rate down</td>
<td>Do not rate down</td>
<td>Do not rate down</td>
<td>High certainty (++++)</td>
<td>Very serious imprecision</td>
<td>Low certainty (++)</td>
<td></td>
</tr>
</tbody>
</table>

Step A1 (necessary for rating the network estimate) Step A2 (optional, not used for rating the network estimate)

---

**Fig 2** | Examples of the process for direct estimate certainty rating. Chronology of rating is depicted from left to right. *Automation available in the spreadsheet provided in supplementary file 1
steps are completed, reviewers obtain a certainty rating which, as above, will inform the assessment of the network estimate. Reviewers can also, if needed (step D), assess imprecision to obtain a final rating of the indirect estimate (steps B1, B2, and B3 in fig 1).

Certainty of the evidence of the most dominant first order loop

The certainty of the evidence of indirect comparisons is based on the certainty ratings of the direct comparisons on which those indirect estimates are calculated (mainly first and second order loops). Therefore, once the certainty ratings are available for all the comparisons that have a direct estimate, indirect estimates certainty ratings can be completed.

**Step B1**

For each indirect comparison, the first step for assessing the certainty of the indirect estimate is to determine which direct comparisons form the most direct of the indirect evidence that connects the interventions of interest (ie, ideally a loop that has only one common comparator between the two interventions being compared, which we term a first order loop) and contributes the most to the indirect estimate. When no first order loop is available (ie, the interventions being compared are not connected through one common comparator), reviewers should then identify a second order loop (ie, two other nodes in the indirect evidence pathway between the two interventions of interest). When more than one alternative is available (multiple first order loops or, eventually, second order loops), reviewers should choose the most dominant in terms of number of studies and participants included.

Supplementary figure 1 from the second update of the living systematic review and NMA on covid-19 prophylaxis illustrates such circumstances. In that situation, for the assessment of the indirect estimate of the comparison between vitamin C and standard of care/placebo on mortality, reviewers had to decide between two first order loops, using hydroxychloroquine or ivermectin as the common comparator. By looking at the size of the nodes and the width of the lines connecting those nodes (which represent the number of patients included in the nodes and number of events for every comparison), reviewers decided that the first order loop through hydroxychloroquine was dominant. Thus, they selected this first order loop to rate the certainty of the indirect estimate certainty.

The initial certainty of the indirect estimate is the lowest of the preliminary certainty ratings of the direct comparisons that constitute the most dominant first (or second) order loops (step B2, fig 1). Figure 3 illustrates the assessment of the indirect estimate for the same comparisons and outcomes.

**Practical considerations**

This process can be facilitated by automation as shown in the supplementary spreadsheet provided (supplementary file 1). Using the spreadsheet, reviewers only need to specify the common comparator in the most dominant first order loop, or the interventions that constitute the second order loop when no common comparator exists, to automatically identify the certainty of the evidence of that loop.

**Assessing intransitivity**

**Step B2**

In addition to the certainty of the most dominant first order loop, the assessment of the certainty of indirect estimate considers intransitivity. For indirect estimates to be valid, the transitivity assumption must not be violated; that is, credible effect modifiers must not be seriously imbalanced between the direct comparisons that contribute to an indirect comparison. The credibility of effect modifiers can be assessed using checklists, or the ICEMAN instrument. When this assumption is not met, because the estimates might reflect an imbalance in effect modifiers instead of a real difference between the interventions of interest, reviewers must rate down the certainty of the evidence.

After the assessment of intransitivity, reviewers obtain a certainty rating of the indirect estimate, which informs the rating of the certainty of the network estimate (indirect preliminary certainty rating omitting imprecision assessment) (step B2 in fig 1).

**Practical considerations**

Intransitivity assessment requires a comparison of potential effect modifiers between the arms that constitute the first order loops of every indirect estimate, which could be facilitated by automation. Using an automated spreadsheet, reviewers could potentially classify every intervention (node) according to presence and magnitude of possible effect modifiers and set rules to automatically obtain intransitivity ratings. Although this functionality is not available in the provided supplementary spreadsheet (supplementary file 1) and reviewers need to manually assess intransitivity, we plan to incorporate it in the future. However, the assessment of intransitivity remains a challenge and the GRADE working group has not yet provided official detailed guidance on how to deal with this domain.

**Final rating of the indirect estimate**

**Step B3**

Similarly, to what was done for the direct estimate, reviewers can finalise the assessment of the indirect estimate by assessing imprecision (step B3 in fig 1). As described for obtaining final rating of the direct estimate (step A2), because imprecision of the indirect estimate does not affect the certainty of the network estimate, obtaining a final certainty rating for indirect estimate is not always necessary (step D); hence, reviewers can choose to skip—or ultimately delay—this step. Figure 3 illustrates the process for the assessment of the indirect estimate of three comparisons on three different outcomes of interest.

**Assessing the certainty of NMA estimates**

The direct and indirect certainty ratings as well as the assessment of incoherence and imprecision inform
<table>
<thead>
<tr>
<th>Comparison</th>
<th>First order loop</th>
<th>Start rating*</th>
<th>Intransitivity assessment</th>
<th>Direct certainty rating to inform the network estimate*</th>
<th>Imprecision assessment*</th>
<th>Final certainty rating of indirect estimate*</th>
</tr>
</thead>
</table>
| Indirect estimate on systemic corticosteroids v standard of care for covid-19. Outcome: Mortality | Systemic corticosteroids ➞ IL6 antagonist ➞ standard of care | Systemic corticosteroids vIL6 antagonist Direct preliminary certainty rating | IL6 antagonist vstandard of care Direct preliminary certainty rating | Lowest certainty of the first loop components | No serious intransitivity | Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty Moderate certainty
ratings (direct preliminary certainty rating to inform NMA (obtained in step A1) or indirect preliminary certainty rating to inform NMA (obtained in step B2)) as a starting point by considering their information size and certainty of evidence. Subsequently, both the assessment of incoherence and imprecision of the network estimate will determine its final certainty rating. Supplementary file 3 presents an algorithm to decide between direct and indirect certainty ratings as a starting point for the NMA certainty rating.

### Practical considerations
The provided supplementary spreadsheet (supplementary file 1) automatically selects between direct and indirect certainty of evidence ratings to inform NMA by following the algorithm presented in supplementary file 3.

### Assessing incoherence
**Step C2**
In addition to deciding whether the NMA estimate rating is based on the direct or indirect estimate rating, the assessment of the certainty of NMA estimate considers incoherence. As both direct and indirect estimates reflect the same effect using different sources of information, they should not differ beyond chance (if that is the case, we refer to the estimates as coherent; if not, they are considered incoherent). When results from direct and indirect estimate diverge beyond what chance can explain and reviewers fail to identify an explanation, reviewers should consider rating down the certainty of the NMA estimate for incoherence.

### Final rating of NMA estimate
**Step C3**
Reviewers should finalise the assessment of the certainty of evidence of NMA estimate by assessing imprecision. As described in step D, this rating informs the best estimate of effect for each comparison. Figure 4 illustrates the process for the assessment of the NMA estimate of three comparisons on three different outcomes of interest.

### Best estimate of effect and certainty rating
**Step D**
At this point, a final certainty rating is available for the NMA estimate. Such a rating might be available for direct and indirect estimates depending on reviewers’ decision to skip imprecision ratings for direct and indirect estimates. Although the purpose of combining direct and indirect estimates is to increase certainty in the estimates of effect, NMA estimates do not always

---

Fig 4 | Examples of certainty ratings for network meta-analysis (NMA) estimates. Chronology of rating is depicted from left to right. *Automation available in the spreadsheet provided in supplementary file 1.
represent the best alternative. For example, when direct and indirect estimates prove incoherent, direct estimates (usually) or indirect estimates (seldom) could have higher certainty than the network estimate, thus constituting a better option to support decision making. Similarly, when the heterogeneity of the network is higher than the heterogeneity of a direct comparison, the direct estimate might be more precise and thus have higher certainty.

When neither of those conditions are met (no incoherence and NMA estimate more precise than direct and indirect estimate), reviewers should consider the NMA estimate with its corresponding certainty of evidence rating as the best. In these cases, assessing imprecision for direct and indirect estimates is not necessary, because reviewers do not use the direct and indirect final ratings. Reviewers might, however, choose to assess imprecision, finalise, and present complete certainty rating of direct and indirect estimates for clarity and transparency.

When, however, direct and indirect estimates are incoherent or NMA estimates are less precise than direct or indirect estimates, reviewers should select the best estimate of effect between the three options (direct, indirect, or NMA). Except for when direct or indirect final ratings result in higher certainty (when the estimate with higher certainty rating should be selected), the NMA estimate represents the best choice (fig 5).

**Practical considerations**

For every possible comparison, reviewers need to select the estimate with higher certainty rating (ie, the final direct, indirect, or NMA estimate), which can also be automatically performed as shown in the supplementary spreadsheet provided (supplementary file 1). In addition, if reviewers should choose to skip steps A2 and B3 and not assess imprecision and finalise direct and indirect certainty of evidence ratings, the spreadsheet automatically informs when these steps are needed to obtain the best estimate of effect with its corresponding certainty of evidence rating.

**Other pragmatic considerations**

**Assessment of imprecision**

Imprecision assessment involves a series of steps that have been previously described in detail including contextualisation, threshold selection, and optimal information size (OIS) assessment. Reviewers should follow published guidance when assessing imprecision in the context of NMA.8

**Practical considerations**

Although imprecision needs to be assessed for every NMA estimate (and eventually every direct and indirect estimate) and every possible comparison, this can be facilitated by rule setting and automation, as shown in the supplementary spreadsheet provided (supplementary file 1). Reviewers can select thresholds
that represent the minimal important effect and a large effect to automatically assess imprecision based on the wideness of the 95% confidence or credible intervals and its relation to the selected thresholds.

Assessing certainty of indirect estimates is not always needed
When the certainty of the direct evidence is high and the contribution of the direct evidence to the network estimate is at least as great as that of the indirect evidence, rating the indirect estimates is not needed. GRADE guidance specifies that the rating of the network estimate is based on the higher of the direct and indirect estimate ratings (supplementary file 3). Therefore, when the direct estimate is high certainty, assessing certainty in the indirect estimate is not needed (ie, it cannot be higher than that of the direct estimate), because reviewers should always choose the direct estimate as the starting point of the NMA certainty rating. However, before skipping the assessment of the certainty of the indirect estimate, reviewers should ensure that the direct estimate is contributing at least as much as the indirect estimate to the NMA estimate. This can be done, for instance, by comparing the widths of the confidence intervals of the direct and indirect estimates or by using statistical tools such as the contribution matrix.

Revising the judgments
After completing the certainty of the evidence assessment process for all relevant comparisons, reviewers need to check the robustness of their decisions. Judgments should have followed the same principles for each comparison when assessing direct, indirect, or NMA estimates. One way to minimise the risk of incoherent judgments within a NMA certainty of the evidence assessment is to set rules as already described in the assessment of imprecision above.

Additional challenges of living reviews
One of the most relevant challenges faced by researchers conducting systematic reviews (particularly for reviews aimed at supporting the development of practice recommendations) is being able to continuously include new information as it becomes available. Keeping systematic review results and conclusions up to date is crucial, especially in contexts where large amounts of information rapidly emerge and policies need to be expeditiously implemented, as is the case of the covid-19 pandemic. The term “living” is currently used to describe a process intended to tackle this issue by continuously searching and incorporating emerging evidence to existing systematic reviews. The process needs a plan that involves all the systematic review stages, from data acquisition to analysis update and results re-interpretation, and can be applied to systematic reviews aiming to perform only pairwise comparisons as well as multiple comparisons with NMA. Performing living systematic reviews with NMA poses additional challenges because for a given clinical scenario, all relevant interventions are incorporated and compared to one another, which can result in a huge amount of information to continuously update and periodically reanalyse. Appropriate result interpretation of living NMAs requires an assessment of the certainty of the evidence for every relevant comparison. Because certainty of the evidence can change every time new information is incorporated to a systematic review, this step becomes a crucial part of the living review process and can introduce an important workload to the review team.

In every update of a living review, computational tools (such as the spreadsheet in supplementary file 1) allows reviewers to differentiate among comparisons that are new (and thus need a full assessment of the certainty of evidence), comparisons for which an important amount of evidence is added (and thus the ratings for each domain of certainty of evidence need to be revised), and comparisons for which little or no new information is available (and thus the certainty of evidence assessment from the previous version of the living review is probably still appropriate and does not need a thorough revision). The team conducting the covid-19 living systematic review and NMA has successfully added and used such features in the spreadsheet (supplementary file 4).

Practical considerations
The process of incorporating new information into a living systematic review can also be facilitated by use of informatic tools for automatic identification of those comparisons where certainty needs to be reassessed and by facilitating access to previous reviewers’ judgments.

Drawing conclusions
The GRADE working group has provided guidance on how to draw conclusion from an NMA using minimally or partially contextualised frameworks. The process is based on a series of rules that allow categorisation of interventions (nodes) according to their effects and certainty of evidence. Interventions placed in higher categories are likely to be more effective than interventions placed in lower categories. Use of these frameworks can be especially useful in the context of large NMAs.

Practical considerations
The process of categorisation of interventions (nodes) according to their effect and certainty of evidence can be automatically performed as shown in the supplementary spreadsheet provided (supplementary file 1). Currently the spreadsheet only offers categorisation based on a minimally contextualised framework, but we plan to incorporate an automated, partially contextualised framework in the future.

Discussion
This article describes the implementation of the GRADE approach for assessing the certainty of evidence from NMA. This assessment presents important challenges, particularly in relation to the exponentially
incremental workload as the number of comparisons in a network increases. We developed the strategy and tools presented in this article based on our experience working on the living systematic review and NMA of randomised trials investigating the prophylaxis and treatment of covid-19,8,11 which led us to generate a process to efficiently assess and reassess the large number of comparisons that were incorporated to the NMA.

To respond to the challenges we faced, we implemented all the strategies to improve efficiency described in published GRADE guidance (table 1) and generated a spreadsheet that automated some of the steps. The strategy and spreadsheet were used by eight reviewers in charge of assessing the certainty of evidence, who provided feedback that led to additional improvements in efficiency. To date, we have assessed the certainty of evidence for the original version and the four updates of the living systematic review and NMA of randomised trials investigating the treatment of covid-19, which included more than 4000 comparisons in the latest version.9 Although the approach has not been formally tested yet, it has been successfully implemented in other systematic reviews with NMA.12-15

Although the presented approach facilitates the assessment of the certainty of the evidence for NMA using GRADE, the process is still complex and requires a substantial workload. We acknowledge that the capabilities of the tool provided as a supplementary spreadsheet can be substantially improved, especially in terms of usability and integration to the evidence based ecosystem.22 For example, the tool could directly import results generated in the data analysis stage, or it could be integrated with platforms for evidence synthesis and guideline development (eg, GRADERPro/GDT and MAGICApp) so that the final GRADE certainty ratings can more easily inform evidence summaries and recommendations. We are working in that direction and expect to provide such a tool in the future. To our knowledge, however, no article so far has described the process of assessing the certainty of evidence from NMAs from a pragmatic perspective; users need to read all the different pieces of GRADE guidance, which is purely conceptual, and are left to creating a process that follows all the guidance. Therefore, we believe that users will benefit from this article describing this process and the existing spreadsheet.

Conclusions

Addressing the certainty of the body of evidence is critical to optimal decision making. The GRADE working group has provided thorough guidance for assessing certainty of evidence from NMA as it involves an assessment of the classic GRADE domains (risk of bias, imprecision, inconsistency, indirectness, publication bias) as well as intransitivity and incoherence, and contemplates not only the certainty in the NMA estimate but also the certainty of the direct and indirect estimates that inform the NMA estimate. Implementing the GRADE approach for NMA is demanding and time consuming but can be substantially facilitated by using a practical stepwise approach as described in this article. By following the proposed approach and incorporating tools that facilitate the tasks, including the supplementary spreadsheet provided, reviewers will achieve a substantial reduction in the NMA certainty of the evidence assessment workload without compromising the rigor.

We thank the authors of the living systematic review and NMA of randomised trials addressing the prophylaxis and treatment of covid-199-11 for their contributions to the development and testing of the presented strategy, Carlos Cuello García for providing a template used for constructing figure 1, and Liang Yao for providing the algorithm presented in supplementary file 3.

In addition to being available as a supplementary material, the spreadsheet to facilitate the assessment of certainty of evidence from network meta-analysis is available at https://www.covid19lnma.com/. We will publish updated versions of this spreadsheet on the website as soon as they become available and will highlight their availability in the homepage of the website.

Contributors: AI and RB-P developed the initial version of the manuscript. AI developed figure 1. GD, DKC, and RAM provided input that resulted important modifications: AI and DKC developed the spreadsheet provided in supplementary materials. RB-P and AI drafted and edited the manuscript, based on feedback from all the authors. All authors approved the final version of the manuscript. AI, who led the project, is the guarantor of this article. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding: None.

Competing interests: All authors have completed the ICMJE uniform disclosure form at www.icmje.org/disclosure-of-interest/ and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in the submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Provenance and peer review: Not commissioned; externally peer reviewed.

Supplementary figure 1: Mortality outcome network plot from the second update of the Prophylaxis Living systematic review and network meta analysis

Supplementary table 1: Process for assessing the certainty of evidence of all comparisons and selecting the best estimate of effect for each comparison in an NMA using an automated spreadsheet

Supplementary file 1: Spreadsheet

Supplementary file 2: Example 1

Supplementary file 3: Algorithm to determine the starting point of NMA certainty

Supplementary file 4: Use of the spreadsheet to facilitate assessments of the CoE during updates of a living systematic review