Journal of Clinical Epidemiology 191 (2026) 112110

ORIGINAL RESEARCH

Paper 2: a semi-automated approach facilitated the assessment of the certainty of evidence for in a network meta-analysis: part 2 − indirect and mixed comparisons

Mohammed Mujaab Kamso[a,b,c,*], Samuel L. Whittle[d,e], Jordi Pardo Pardo[f], Rachelle Buchbinder[d], George Wells[g,h], Rob Deardon[i,j], Tolulope Sajobi[a], George Tomlinson[k], Jesse Elliott[h], Jocelyn Thomas[a,b,c], Shannon E. Kelly[g,k], Romina Brignardello-Petersen[l], Glen S. Hazlewood[a,b,c]

[a]*Department of Community Health Sciences, Cumming School of Medicine, University of Calgary, Calgary, Canada*
[b]*Arthritis Research Canada, University of British Columbia, Vancouver, Canada*
[c]*McCaig Institute for Bone and Joint Health, University of Calgary, Calgary, Canada*
[d]*Musculoskeletal Health and Wiser Health Care Units, School of Public Health and Preventive Medicine, Monash University, Melbourne, Australia*
[e]*Rheumatology Department, The Queen Elizabeth Hospital, Adelaide, Australia*
[f]*Centre for Practice-Changing Research, Ottawa Hospital Research Institute, The Ottawa Hospital - General Campus, Ottawa, Canada*
[g]*School of Epidemiology and Public Health, University of Ottawa, Ottawa, Canada*
[h]*Cardiovascular Research Methods Centre, University of Ottawa Heart Institute, Ottawa, Canada*
[i]*Faculty of Science, Department of Mathematics & Statistics, University of Calgary, Calgary, Canada*
[j]*Faculty of Veterinary Medicine, University of Calgary, Calgary, Canada*
[k]*Department of Medicine, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada*
[l]*Department of Health Research Methods, Evidence, and Impact, McMaster University, Hamilton, Canada*

Accepted 10 December 2025; Published online 16 December 2025

Abstract

Objectives: To implement a semiautomated approach to facilitate rating the Grading, Recommendation, Assessment, Development and Evaluation certainty of evidence (CoE) for indirect and network meta-analysis (NMA) estimates.

Methods: We developed and implemented algorithms for generating automated ratings for the CoE for indirect and network estimates in two living NMAs of rheumatoid arthritis treatment. At the indirect stage, inputs included CoE ratings for direct estimates and the contribution matrix. Intransitivity ratings were assigned based on the indirectness ratings of the two direct estimates with the highest percent contribution. An online tool (customized to our project) facilitated assessment of imprecision on the network estimate. Automated ratings were reviewed by two independent experts.

Results: Across 1306 indirect comparisons, the contribution matrix identified the dominant branches of evidence regardless of whether a single first order loop was present (80%) or not. The reviewers agreed with all automated CoE ratings for incoherence ($n = 34$), network estimates ($n = 34$) and imprecision ($n = 1447$). They agreed with the automated intransitivity algorithm except when the total contribution of the top-two direct estimates was low (eg, <50%, which occurred in 38% of the estimates).

Conclusion: Automated approaches facilitated CoE ratings for indirect and network estimates. Further work is required to define appropriate algorithms for intransitivity.

Keywords: Randomized controlled trials (RCTs); Network meta-analysis; Bayesian; Rheumatoid arthritis; GRADE; Living systematic reviews; Semiautomation

Funding: This work was supported by the Eyes High Doctoral Recruitment Scholarship and the Canadian Institutes for Health Research (CIHR) [Funding Reference Number (FRN) 178375].

* Corresponding author. Departments of Medicine and Community Health Sciences, Cumming School of Medicine, University of Calgary, 3280 Hospital Drive NW, Calgary, T2N 4Z6, Canada.

E-mail address: mohammedmujaab.kamso@ucalgary.ca (M.M. Kamso).

https://doi.org/10.1016/j.jclinepi.2025.112110

What is new?

Key findings

- We developed a semi-automated approach to rate certainty of indirect and network meta-analysis estimates.

What this adds to what is known?

- Algorithms can assign most certainty of evidence (CoE) ratings for indirect and network estimates using Grading of Recommendations Assessment, Development and Evaluation.
- Intransitivity judgements from direct evidence indirectness are novel but need refinement.

What is the implication and what should change now?

- Semi-automating CoE in living network meta-analysis can reduce workload and improve transparency over time.
- The online imprecision tool is viewable but currently customized to our project.

1. Background

In a network meta-analysis (NMA), many judgments need to be made when evaluating the certainty of evidence (CoE). In the Grading of Recommendations Assessment, Development and Evaluation (GRADE) approach, the CoE is rated at the direct, indirect, and network evidence phases separately [1,2]. Therefore, in extensive networks with numerous interventions, there is a large workload associated with evaluating the CoE across the various phases of NMA results. In the context of a living NMA, this challenge is further amplified due to the dynamic nature of the network geometry and the continuous influx of new evidence.

In the GRADE approach, after rating the certainty of direct evidence, indirect evidence is evaluated [1−3]. Indirect estimates are generated through models like ‘‘node-splitting’’ [1,4], where trials directly comparing the interventions are removed and the NMA is reanalyzed. When assessing the certainty of indirect evidence, authors first consider the CoE of the direct estimates that contribute most to the indirect estimate and the potential for intransitivity. This is often performed by reviewing a network graph and focusing on the first-order loops of evidence [2,3,5]. Intransitivity is assessed by considering differences in effect modifiers across the major evidence branches [3]. Although efforts to evaluate intransitivity in NMAs have been made, a formalized implementation is lacking [3], leading to many NMAs omitting or minimally detailing the intransitivity assessment [6].

After rating direct and indirect evidence, review authors rate the certainty of the network estimate for each comparison [1]. This is straightforward when only direct or indirect evidence is present, as the NMA estimate's (before assessing imprecision) initial rating is the same as the direct or indirect estimate. When both types of evidence are available, GRADE guidance suggests starting with the certainty rating of the evidence contributing most to the NMA estimate, based on the width of the CIs or credible intervals. Alternatively, the contribution matrix can be used to determine the primary evidence source [1]. Authors then assess coherence between direct and indirect evidence, often using statistical tests and CI or credible interval overlap [7]. Finally, after considering potential downgrades for incoherence, a final CoE rating is assigned, factoring in imprecision and outcome thresholds relevant to the framework used (minimally or partially contextualized framework) [8,9].

In part 1, we described and implemented an approach for generating semiautomated judgments for rating the certainty of the direct evidence within an NMA [10]. Here, in part 2, we extend this approach to rating of the indirect and NMA estimates. Again, we use the term ‘‘semiautomated’’ to mean an approach in which suggested certainty ratings are automatically generated based on the prespecified criteria (which may be specific to a review), then reviewed by experts, who may modify the ratings, while providing a rationale. The aims of developing this approach were to improve the efficiency, transparency, and reproducibility of rating the CoE within a living NMA, where many judgemnts need to be made and updated over time. This paper describes the approach we developed and its application in two living NMAs.

2. Methods

2.1. Description of source NMA

An outline of our approach, illustrating how each step aligns with the overall GRADE-NMA process, is provided in Figure 1. To develop and test our approach, we used data from an ongoing living NMA of rheumatoid arthritis drug therapy randomized trials analyzed in the Bayesian paradigm [11], that informs living guidelines in Canada and Australia [12,13]. The two specific NMAs compare disease-modifying antirheumatic drug (DMARD) options at two time points: initial therapy (DMARD-naïve), and after an inadequate response to antitumor necrosis factor (TNF) therapy (TNF-inadequate responder). In this paper, we focus on the three critical outcomes in the review: American College of Rheumatology response criteria-50, a dichotomous outcome of treatment response; radiographic progression, a continuous outcome; withdrawals due to adverse events, assessed as a rate [11]. The statistical

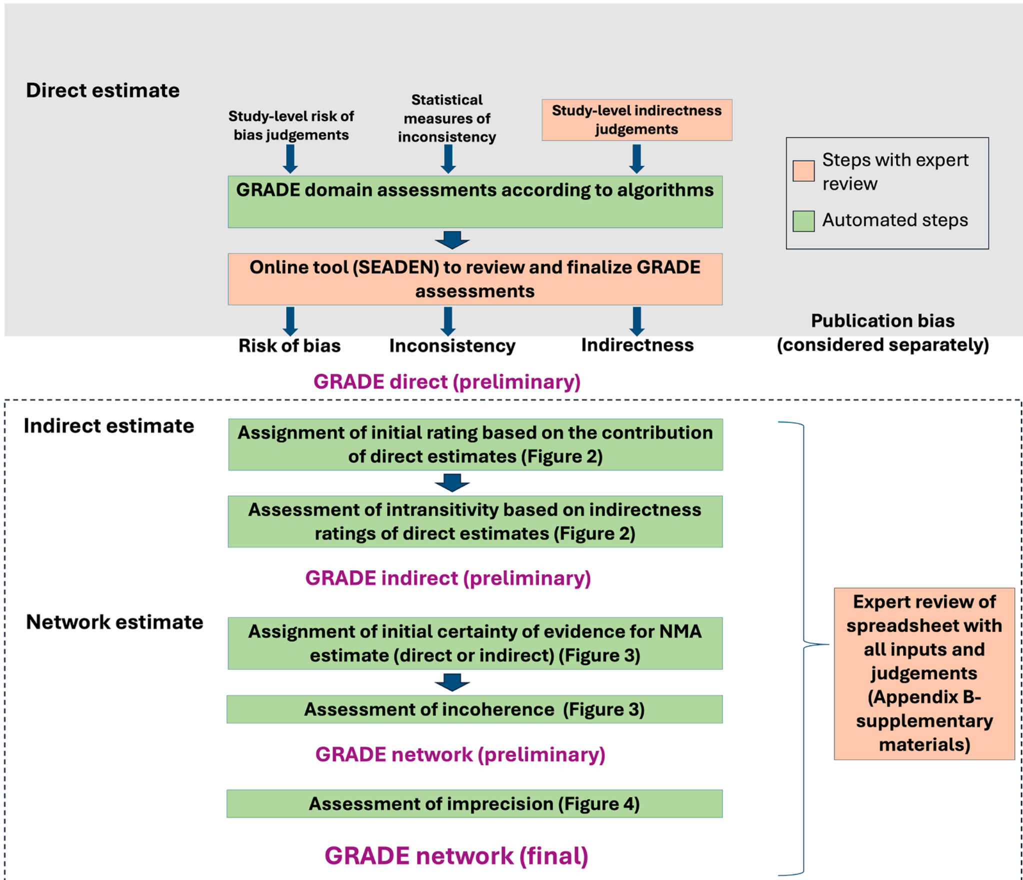

Figure 1. Overview of our approach within the overall GRADE-NMA process. This paper focuses on generating the certainty rating for the indirect and mixed estimate. GRADE, Grading of Recommendations Assessment, Development and Evaluation; NMA, network meta-analysis.

model and software use is described in Appendix A. We have previously assessed the GRADE CoE for each of these direct estimates, also incorporating semiautomated approaches to facilitate the ratings [10].

2.2. *Certainty ratings for indirect evidence*

2.2.1. *Initial rating*

Consistent with GRADE guidance [1−3], indirect estimates were assigned the lowest rating from the two direct estimates contributing most to the indirect evidence. The percentage contribution of each direct estimate was calculated using the contribution matrix for the NMA estimate, which is a frequentist approach, but can be used as an approximation for a Bayesian NMA, where no similar approach exists [1,14]. We used the R package *netmeta*, scaling the values after removing the direct evidence contribution. The contribution matrix is used to evaluate how much each direct evidence (from head-to-head trials) in terms of percentage or proportion contributes to the relative treatment effect estimates in the entire NMA. For instance, if the contribution for an NMA estimate A−C was 50% from direct evidence, and 25% each from two branches of indirect evidence (A−B, B-C), A−B and B−C each contributed 50% to the indirect estimate for A−C. We focused on the top two direct estimates to align with GRADE guidance to typically focus on the dominant first order loop, which usually contributes the most to the indirect evidence [1,2]. We compared the similarity of branches identified using our automated approach with a manual approach, where one reviewer (K.M.) manually identified the first-order loop(s) by inspecting the network plot.

2.2.2. *Intransitivity and final rating*

To inform the ratings of intransitivity, we used the indirectness ratings for the two main branches of direct evidence identified in the first step. Following GRADE guidance [1−3], each direct estimate received an indirectness rating based on judgments regarding population, intervention, comparator, and outcome (PICO) characteristics in the underlying trials. Ratings included ‘‘serious’’, ‘‘not serious’’ (standard GRADE categories), and an additional ‘‘some concerns’’ category for added nuance in automated ratings. Through group consensus, we developed rules for assigning automated intransitivity ratings based on these indirectness ratings, outlined in Figure 2. If either or both direct estimates were rated serious for indirectness, we rated intransitivity as ‘‘serious’’ and downgraded the CoE by one level. If both were rated "not serious," the intransitivity rating was "not serious," and not downgraded. If one or both branches had "some concerns" (without being serious), intransitivity was flagged as "some concerns" for expert review, but was not automatically downgraded. We recognized that GRADE advises not downgrading for intransitivity if indirectness concerns (extent or direction of suspected effect modification) in the branches of direct evidence are the same, as this could be considered double-downgrading. However, in practice, this is challenging to evaluate, and we chose to be conservative, erring on the side of rating down for intransitivity if concerns were present.

2.3. Certainty rating for network estimates

2.3.1. Initial rating

Following GRADE guidance [1−3], the initial automated rating for the network estimate was the rating for the estimate (direct or indirect) that contributed the most to the network estimate [1,2,7], unless the contribution was similar, in which case GRADE guidance is to choose the higher of the two certainty ratings. We used the contribution matrix to operationalize this, choosing a threshold of >60% to indicate an estimate contributed the most to the network estimate and 40%−60% to indicate a similar contribution (Fig 3).

2.3.2. Incoherence

To assess incoherence, we performed an incoherence test comparing direct and indirect estimates under the null hypothesis of equality, using the Bayesian approach (node-split) for local incoherence examination [2,7,8,15]. Following GRADE guidance [1−3], when the direct and indirect evidence contributed similarly to the network estimate (40%-60%) and the *P* value was low, we downgraded for incoherence, choosing the commonly used threshold of *P* value <.05 (Fig 3). If the *P* value was between 0.05 and 0.20, we assigned the judgment of "some concerns" to flag for experts to review and decide whether it was appropriate to downgrade. When one estimate (direct or indirect) dominated in its contribution to the direct evidence (>60%), we did not downgrade for incoherence, as per GRADE guidance [1−3]. Rather, we flagged a *P* value <.05 as "some concerns" for expert review (Fig 3).

2.4. Imprecision and final certainty rating

The final step was assigning a judgment for imprecision on the NMA estimate. Using a minimally contextualized framework, we rated imprecision based on absolute treatment effects, calculated from relative effects and the median-pooled response across placebo (reference) arms as the assumed comparator risk [8,9]. We defined thresholds for minimally important and large effects for each outcome (Table 1), and then followed GRADE guidance

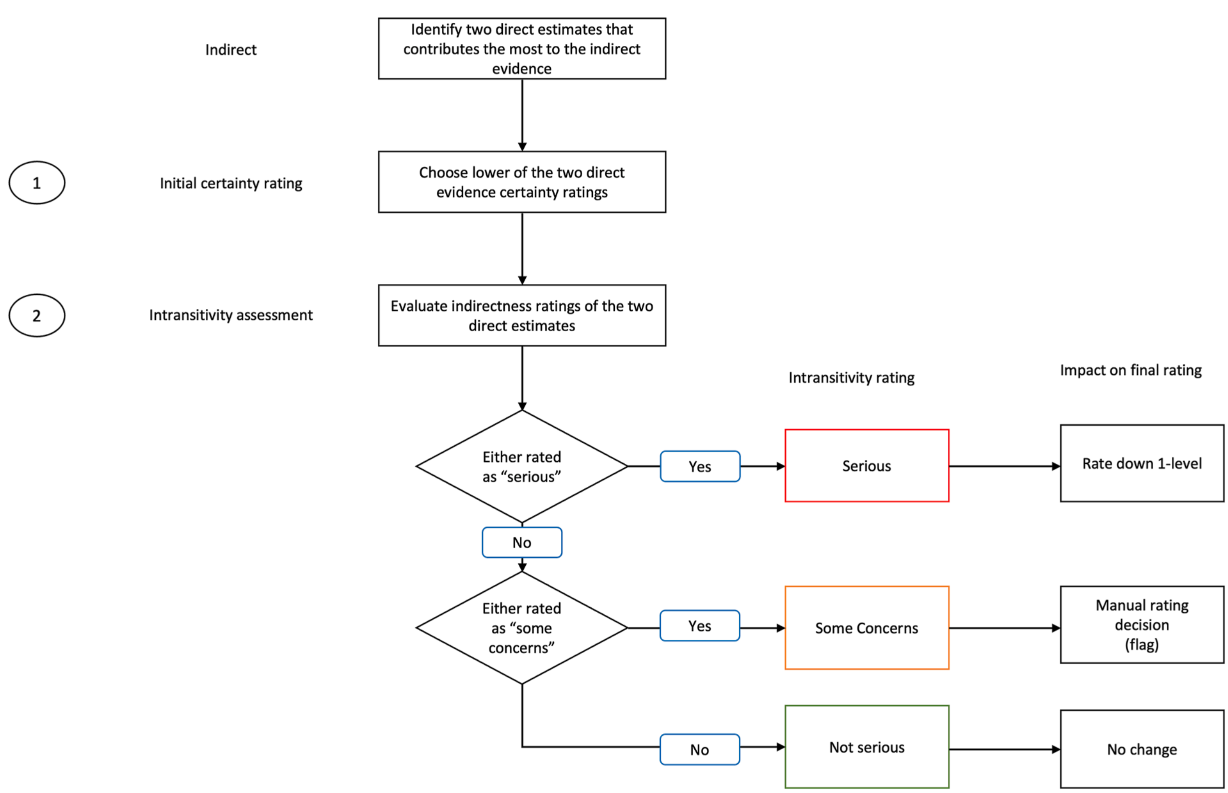

Figure 2. Algorithm for assigning certainty of evidence (CoE) ratings for the indirect estimates.

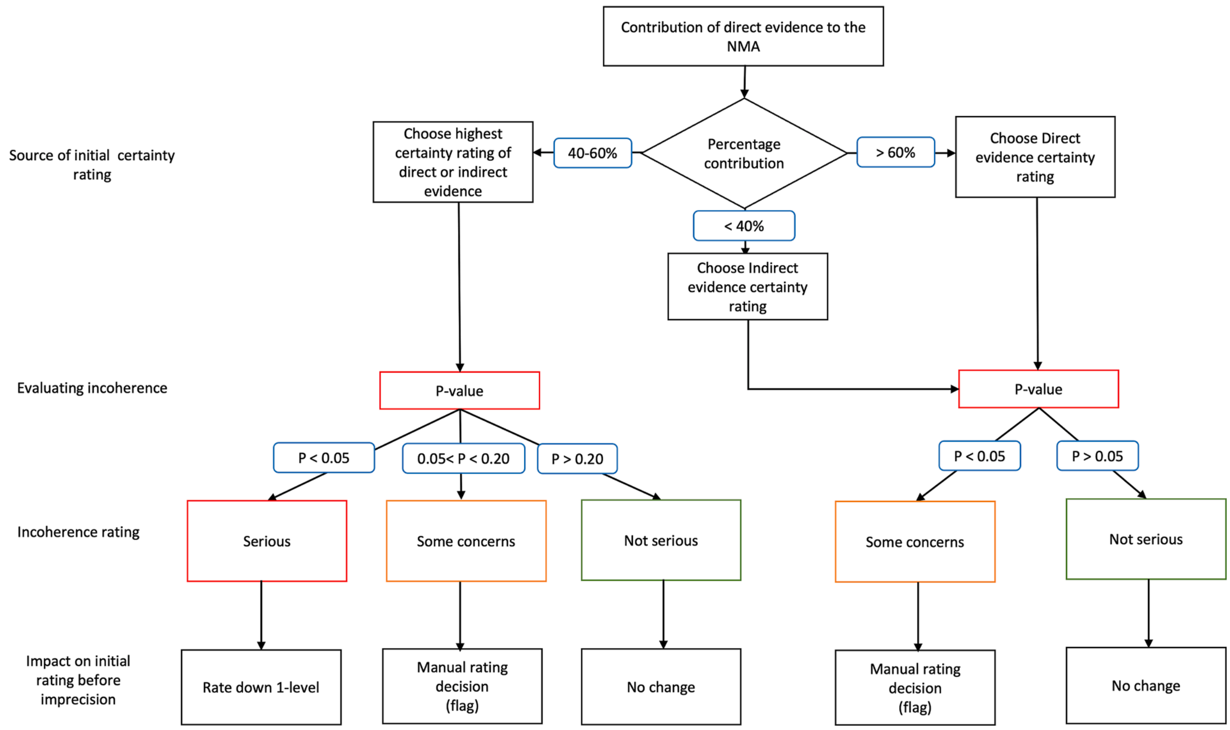

Figure 3. Algorithm for assigning certainty of evidence (CoE) ratings for the NMA estimates before imprecision assessment. NMA, network meta-analysis.

[9] to rate down by one, two, or three levels based on the width of the CI or credible interval of the absolute effect (summarized in Fig 1, Appendix A). We created an online tool to allow users to adjust values for minimally important and large effects, and the chosen threshold (null or minimally important effect) for the certainty ratings (Fig 4, with the tool available at [imprecision application]). Certainty ratings are updated automatically, and users can download the results. The final rating included an "extremely low" category to better reflect the number of times the evidence was downgraded, but this can be collapsed to "very low".

2.5. *Expert confirmation of judgments*

After assigning all the judgments, we generated a dataset (Excel file) that included all judgments and the relevant data that informed these judgments (eg, the P value for incoherence, imprecision thresholds). Two reviewers (G.H. and J.P.) independently reviewed this dataset and all judgments for accuracy. They also had access to all other data in the review (eg, forest plots). After review, they discussed and proposed modifications to the algorithms. Given the large volume of judgments and the fact that most judgments follow directly from the algorithms, we did not formally assess agreement between reviewers. Rather, we sought to identify, through independent expert review, any issues with the chosen algorithms to suggest improvements.

3. Results

Across the three outcomes of the two living NMAs, there were 1447 NMA estimates (Table 2). These were informed by direct evidence only 141 times (9.7%), indirect evidence only 1272 times (87.9%), and both direct and indirect evidence 34 times (2.3%).

3.1. *Indirect estimates*

3.1.1. *Network geometry (manual review)*

Of the 1306 indirect estimates, 1054 (80.7%) had a first order loop, with four of these having multiple pathways for a first-order loop. Of the 252 indirect estimates without the first-order loops, 86 had multiple pathways for higher order loops. The network diagram presented in Figure 5 shows an example of a comparison where there was a first-order loop and multiple pathways (two) for a second-order loop.

3.1.2. *Comparison between contribution matrix and manual review*

The automated approach of choosing the highest two contributing direct estimates generally aligned a manual review of the network plot, but offered advantages in always selecting the dominant evidence, regardless of network geometry. When a single first order loop was present, the two direct estimates with the highest percent contribution were from this first-order loop in 90.6% of cases (951/1050). When multiple first order loops were present (n = 4), the contribution matrix successfully identified the dominant

Table 1. Chosen thresholds for rating the certainty of evidence (CoE) for imprecision following the minimally contextualized framework

Outcomes	Small effect (MID)	Large effect threshold
ACR50	5%	20%
Withdrawals due to adverse events	5%	20%
Radiographic progression (SvDH)	5 points	10 points

ACR50, American College of Rheumatology response criteria-50; SvDH, Sharpe van der Heijde.

For the continuous outcome, the minimally important threshold was the minimally important difference (MID), based on published values [16]. For dichotomous and rate outcomes, the thresholds were agreed upon by consensus, as were the large effect thresholds.

loop, which contributed between 45% and 75% to the indirect estimate. In the 99 cases without a first-order loop, the two highest contributing estimates came from second-order loops (n = 97), or third (n = 1) and fourth-order loops (n = 1). When the highest order loop was second (n = 157) or third (n = 8), the contribution matrix always identified direct estimates from those loops. However, despite always selecting the dominant branches, the top two direct estimates accounted for >50% of the total evidence for only 62% (811/1306) of the indirect estimates (Fig 6).

3.1.3. Intransitivity

A total of 1306 indirect evidence assessments were evaluated for intransitivity. Of these, 620 (47.5%) were flagged as "not serious", 359 (27.5%) as "some concerns", and 327 (25.0%) as "serious" and rated down. The reviewers generally agreed with the algorithm-assigned judgments but were less confident when the percentage contribution of the indirect evidence considered (top two branches) was low (eg, especially <50%). For the estimates rated down for intransitivity, 302/327 (92.4%) had serious

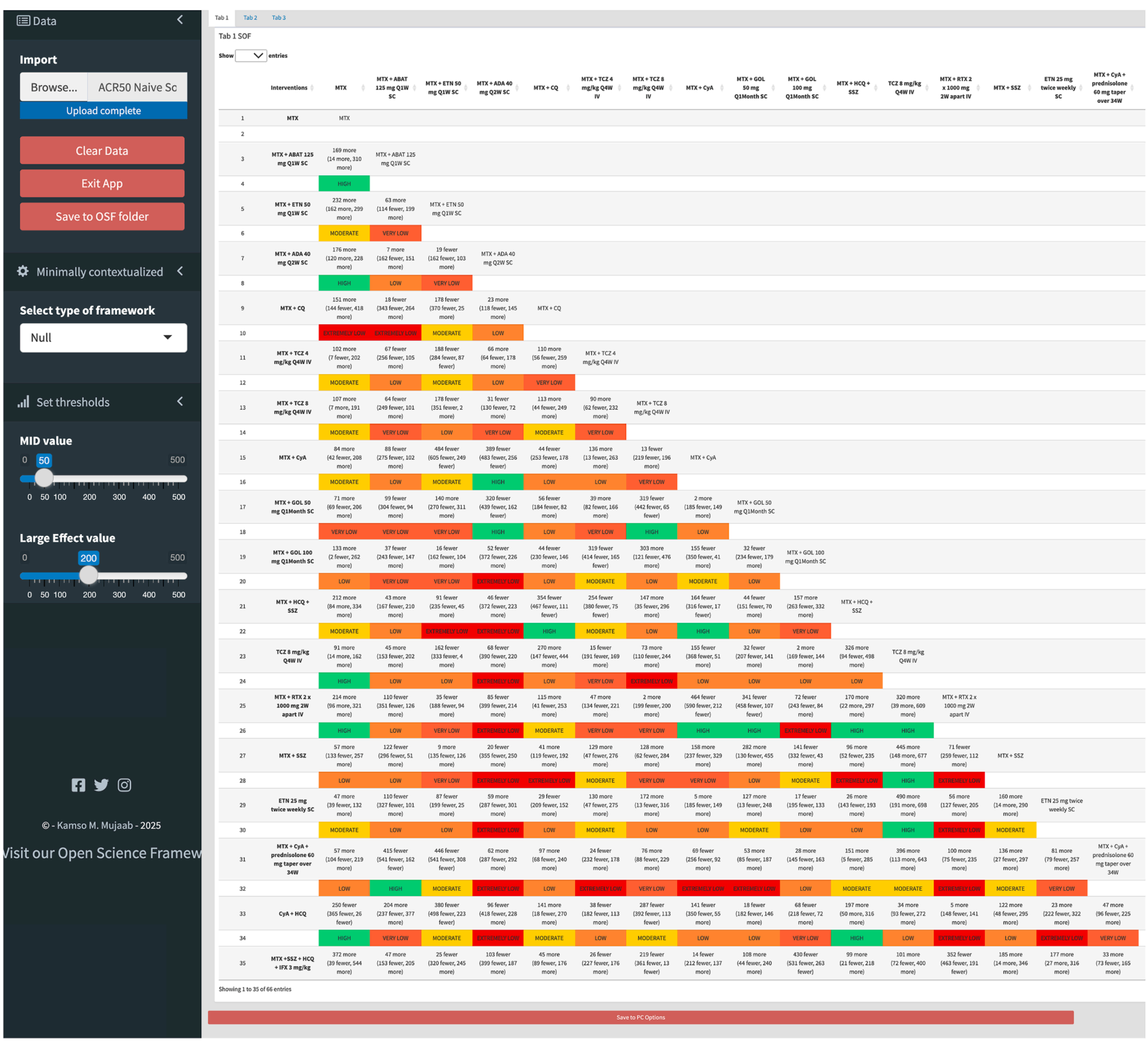

Figure 4. Screenshot of TAB1 showing the various comparisons, risk difference (CrI) and NMA GRADE rating. CrI, credible interval; GRADE, Grading of Recommendations Assessment, Development and Evaluation; MID, minimal important difference; NMA, network meta-analysis; SOF, summary of findings.

Table 2. Distribution of direct, indirect, and network evidence across the various outcomes

Review	Outcome	Outcome type	Number of estimates: Direct evidence only	Indirect evidence only	Both direct and indirect evidence	Total
TNF-IR	ACR50	Dichotomous	13	62	3	78
	Withdrawals due to adverse events	Rate	12	43	-	55
	Radiographic progression	Continuous	-	-	-	-
DMARD-naive	ACR50	Dichotomous	38	476	14	528
	Withdrawals due to adverse events	Rate	39	414	12	465
	Radiographic progression (final values)	Continuous	4	14	3	21
	Radiographic progression (change from baseline)	Continuous	35	263	2	300
	Total		141	1272	34	1447

ACR50, American College of Rheumatology response criteria-50; DMARD-naïve, disease-modifying antirheumatic drug; TNF-IR, tumor necrosis factor—inadequate responder.

indirectness concerns in just one branch of direct evidence, and 25 (7.6%) had concerns in both. In the vast majority of cases, rating down for intransitivity had little meaningful impact on the final NMA estimate. Of the 327 estimates, the final NMA estimate would have been low certainty or worse in 312 (95.4%) even if the indirect estimate was not rated down for intransitivity. In the 25 estimates where both branches had indirectness concerns (and it could therefore be considered inappropriate to downgrade for both indirectness and intransitivity), the final NMA estimate had a certainty rating of very low or extremely low 22/25 times (low for the remaining three estimates), suggesting there was little practical impact in our reviews if we had chosen not to downgrade both indirectness in this situation.

3.2. Network estimates

Of the 34 comparisons with both direct and indirect estimates across the outcomes examined, 24 (71%) had either the direct or indirect estimate contributing over 60% to the network estimate. The direct evidence dominated for thirteen estimates, and the indirect evidence dominated for the other eleven. Ten estimates had a similar contribution from the direct and indirect evidence (40%—60%) to the network estimate. The reviewers were comfortable with the judgments and did not propose any modifications.

3.2.1. Incoherence

Of the 34 incoherence assessments, 26 (76%) were rated as "not serious" and 8 (24%) were flagged as "some concerns" with P values ranging from 0.10 to 0.86. After examining the scenarios flagged as "some concerns" the reviewers decided not to further rate them down for incoherence, suggesting that our algorithm could be modified to only flag P values between 0.05 and 0.10.

3.2.2. Imprecision

Of the 1447 network estimates, 1176 (81%) were rated down for imprecision: "serious" 505 times (43%), "very serious" 579 times (49%), and "extremely serious" 92

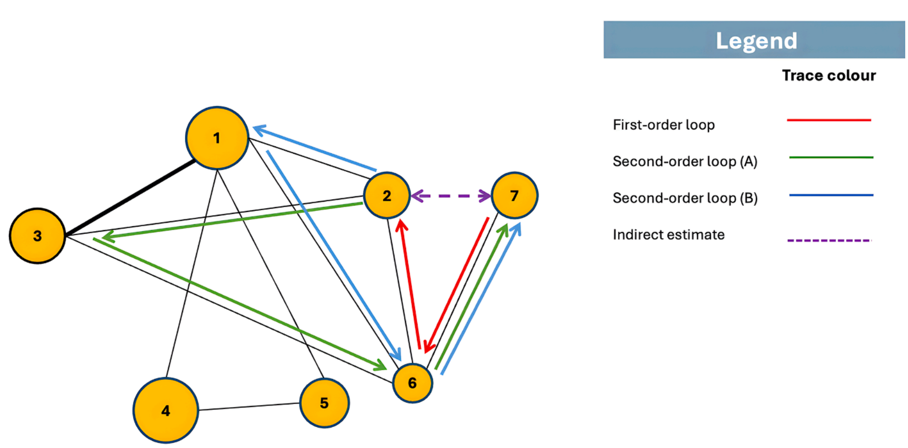

Figure 5. Example illustrating loops of evidence relating to intervention 2 and 7.

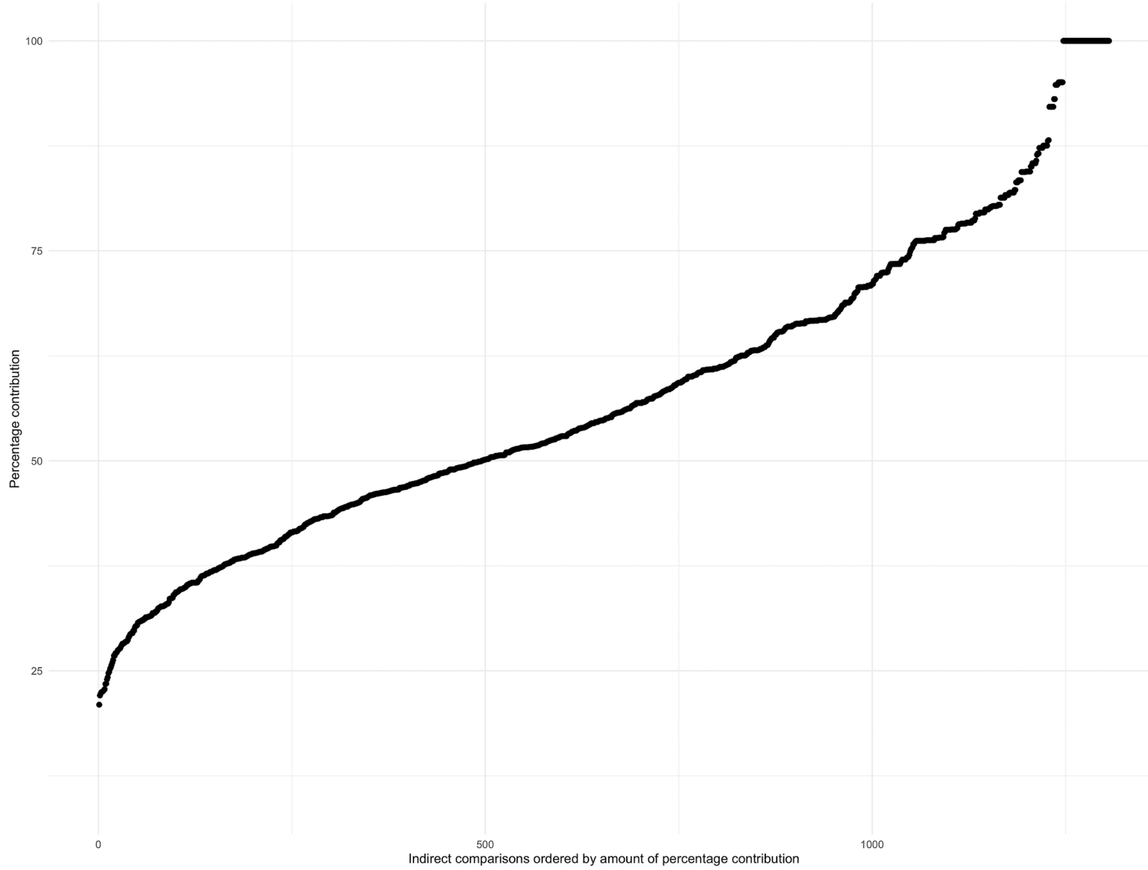

Figure 6. Total percentage contribution of highest two contributing branches of direct evidence for each indirect estimate.

times (8%), whereas 271 (19%) were not rated down. Reviewers were comfortable with the final judgments and did not make any changes. A link to an excel sheet with all ratings and inputs for the automated judgments is available in Appendix B.

4. Discussion

In this study, we found a semiautomated approach facilitated rating of the CoE for indirect and network estimates within an NMA. The contribution matrix was useful in the algorithms for both the indirect and NMA estimates, as it allowed us to quantify the contribution of the source evidence for each. We introduced a novel approach for using ratings of indirectness of the direct estimates to inform judgments of intransitivity, and developed a tool that automatically assigned certainty ratings for imprecision based on assigned thresholds for small and large effects. Our approach does not eliminate or reduce the need for human judgment in assigning the CoE, but rather uses human judgment to develop the rules and algorithms following GRADE guidance [1−3], then applies these consistently across the NMA. Taken together, this study offers an approach with the potential to save time, enhance transparency, and ensure consistency in CoE evaluation within an NMA.

We used indirectness ratings of direct estimates to inform intransitivity judgments, which make sense conceptually, as both indirectness and intransitivity relate to confidence in effect modification. However, if important effect modification is suspected in both branches and the modifiers are the same, one would not rate down for intransitivity [3]. In practice though, this is difficult to assess. We chose to rate down for intransitivity if either or both branches had indirectness concerns, regardless of whether they involved the same PICO elements. Although this could be seen as inappropriately double-downgrading, we preferred a conservative approach, although acknowledge that this decision is subjective, and other reviewers might adopt a different threshold. This situation was rare though and had little practical impact on the certainty rating of the final estimate.

The choice of using only the top two branches of evidence in the intransitivity algorithm also likely needs modification. This is, however, not necessarily a critique of our approach, but rather a broader issue with intransitivity assessments in complex networks, where focusing only on the first-order loops or the dominant two branches may not capture the majority of the contributing evidence. This requires further research, comparing thresholds with expert review and extending the approach to other NMAs.

Salanti et al [14] proposed an alternative approach to rating CoE in NMA evidence, developed into the Confidence In Network meta-analysis (CINeMA) web tool. Our approach integrates concepts from CINeMA, namely borrowing the use of the contribution matrix [2,14]. Our paper makes no assessment regarding the relative merits of each approach. Rather, we propose a potential approach to facilitate assigning GRADE CoE judgments, when authors have chosen to use this approach. There could be potential value in comparing our approach to the CINeMA approach to determine the impact on the certainty rating. A recent study compared CoE ratings between the CINeMA and GRADE Working Group approaches. Specifically, it found that the GRADE Working Group approach generally assigns a higher CoE compared to the CINeMA framework due to differences on how both evaluate inconsistency and the indirect evidence [17]. Our semiautomated approach could help improve the efficiency of assigning GRADE CoE judgments, while also ensuring consistency and transparency of the judgments over time in a living NMA context. Other groups have identified the large workload involved in assessing the GRADE CoE in an NMA as a practical barrier [1,5,18,19]. A recent publication proposed practical ways to automate some of the steps by implementing rules in a spreadsheet [5].

A potential risk with our approach is that the automated judgments could diminish thoughtful assessment by the reviewers. By developing algorithms and having experts review the automated ratings and document any changes they make, end users can look to a small number of algorithms and modified ratings when determining whether they agree with the certainty ratings. This approach can be applied to networks of varying types and scales, as evidenced by the breadth of outcomes assessed. Although the online tools are customized to our project, the algorithms and overall approach can be used without them, and may provide benefits to NMA authors. For instance, the algorithm-assigned judgments could be reviewed and confirmed by experts outside of the online tools.

In addition to further research evaluating our approach, future work could merge the web tools for evaluating the CoE of the direct, indirect, and network estimates from an NMA model, and generalize them so they can be broadly applied to any NMA. Finally, a valuable next step would be the implementation of the optimal information size at the NMA phase when evaluating the imprecision [9,20,21].

5. Conclusion

A semiautomated approach with decision rules can facilitate the evaluation of the CoE from a living NMA. This could result in a significant increase in efficiency, while also aiding in maintaining transparency and consistency throughout the process.

CRediT authorship contribution statement

Mohammed Mujaab Kamso: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Samuel L. Whittle:** Writing – review & editing, Resources, Conceptualization. **Jordi Pardo Pardo:** Writing – review & editing, Investigation, Conceptualization. **Rachelle Buchbinder:** Writing – review & editing, Resources, Funding acquisition, Conceptualization. **George Wells:** Writing – review & editing, Resources, Funding acquisition, Conceptualization. **Rob Deardon:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Tolulope Sajobi:** Writing – review & editing, Methodology, Conceptualization. **George Tomlinson:** Writing – review & editing, Methodology, Conceptualization. **Jesse Elliott:** Writing – review & editing, Resources, Conceptualization. **Jocelyn Thomas:** Writing – review & editing, Resources, Conceptualization. **Shannon E. Kelly:** Writing – review & editing, Resources, Methodology, Conceptualization. **Romina Brignardello-Petersen:** Writing – review & editing, Conceptualization. **Glen S. Hazlewood:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Methodology, Investigation, Funding acquisition, Conceptualization.

Declaration of competing interest

S.W. is supported by an Australia and New Zealand Musculoskeletal (ANZMUSC) Clinical Trial Network Practitioner Fellowship and by a grant from The Hospital Research Foundation Group. R.B. is supported by an Australian National Health and Medical Research Council (NHMRC) Investigator Fellowship (APP1194483).

Acknowledgments

The authors would like to express their sincere gratitude to Renea Johnston and Vanessa Glennon for their valuable contributions to this work.

Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.jclinepi.2025.112110.

Data availability

Data will be made available on request.

References

[1] Brignardello-Petersen R, Bonner A, Alexander PE, Siemieniuk RA, Furukawa TA, Rochwerg B, et al. Advances in the GRADE approach to rate the certainty in estimates from a network meta-analysis. J Clin Epidemiol 2018;93:36−44.
[2] Puhan MA, Schünemann HJ, Murad MH, Li T, Brignardello-Petersen R, Singh JA, et al. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. BMJ 2014;349:g5630.
[3] Brignardello-Petersen R, Tomlinson G, Florez I, Rind DM, Chu D, Morgan R, et al. Grading of recommendations assessment, development, and evaluation concept article 5: addressing intransitivity in a network meta-analysis. J Clin Epidemiol 2023;160:151−9.
[4] van Valkenhoef G, Dias S, Ades AE, Welton NJ. Automated generation of node-splitting models for assessment of inconsistency in network meta-analysis. Res Synth Methods 2016;7:80−93.
[5] Izcovich A, Chu DK, Mustafa RA, Guyatt G, Brignardello-Petersen R. A guide and pragmatic considerations for applying GRADE to network meta-analysis. BMJ 2023;381:e074495.
[6] Wang Y, Xia R, Pericic TP, Bekkering GE, Hou L, Bala MM, et al. How do network meta-analyses address intransitivity when assessing certainty of evidence: a systematic survey. BMJ open 2023;13:e075212.
[7] Brignardello-Petersen R, Mustafa RA, Siemieniuk RA, Murad MH, Agoritsas T, Izcovich A, et al. GRADE approach to rate the certainty from a network meta-analysis: addressing incoherence. J Clin Epidemiol 2019;108:77−85.
[8] Brignardello-Petersen R, Florez ID, Izcovich A, Santesso N, Hazlewood G, Alhazanni W, et al. GRADE approach to drawing conclusions from a network meta-analysis using a minimally contextualised framework. BMJ 2020;371:m3900.
[9] Zeng L, Brignardello-Petersen R, Hultcrantz M, Mustafa RA, Murad MH, Iorio A, et al. GRADE Guidelines 34: updated GRADE guidance for imprecision rating using a minimally contextualized approach. J Clin Epidemiol 2022;150:216−24.
[10] Kamso MM, Whittle SL, Pardo JP, Buchbinder R, Wells G, Deardon R, et al. A semi-automated approach facilitated the assessment of the certainty of evidence in a network meta-analysis: part 1−direct comparisons. J Clin Epidemiol 2025:112109.
[11] Hazlewood GS, Whittle SL, Kamso MM, Akl EA, Wells GA, Tugwell P, et al. Disease-modifying anti-rheumatic drugs for rheumatoid arthritis: a systematic review and network meta-analysis. Cochrane Database Syst Rev 2020;2020:CD013562.
[12] (ANZMUSC) ANZMCTN. Australia & New Zealand Musculoskeletal Clinical Trials Network (ANZMUSC). An Australian living guideline for the pharmacological management of inflammatory arthritis, Making GRADE the Irresistible Choice (MAGICapp) - Guidelines and Evidence Summaries. Available at: https://app.magicapp.org/#/guideline/LqRV3n2023. Accessed January 6, 2025.
[13] Hazlewood GS, Pardo JP, Barnabe C, Schieir O, Barber CE, Proulx L, et al. Canadian Rheumatology Association living guidelines for the pharmacological management of rheumatoid arthritis with disease-modifying antirheumatic drugs. J Rheumatol 2022;49:1092−9.
[14] Salanti G, Del Giovane C, Chaimani A. Evaluating the quality of evidence from a network meta-analysis. PLoS One 2014;9(7):e99682.
[15] van Valkenhoef G, Tervonen T, de Brock B, Hillege H. Algorithmic parameterization of mixed treatment comparisons. Stat Comput 2012;22:1099−111.
[16] Bruynesteyn K, van der Heijde D, Boers M, Saudan A, Peloso P, Paulus H, et al. Determination of the minimal clinically important difference in rheumatoid arthritis joint damage of the Sharp/van der Heijde and Larsen/Scott scoring methods by clinical experts and comparison with the smallest detectable difference. Arthritis Rheum 2002;46:913−20.
[17] Noori A, Sadeghirad B, Thabane L, Bhandari M, Guyatt GH, Busse JW. The GRADE Working Group and CINeMA approaches provided inconsistent certainty of evidence ratings for a network meta-analysis of opioids for chronic noncancer pain. J Clin Epidemiol 2024;169:111276.
[18] Gartlehner G, Nussbaumer-Streit B, Devane D, Kahwati L, Viswanathan M, King VJ, et al. Rapid reviews methods series: guidance on assessing the certainty of evidence. BMJ Evid Based Med 2024;29:50−4.
[19] Andrić A, Peričić TP, Tokalić R, Marušić A, Puljak L. Experiences and challenges faced by systematic review authors in using the GRADE approach: a qualitative study. J Clin Epidemiol 2025;185:111870.
[20] Brignardello-Petersen R, Guyatt GH, Mustafa RA, Chu DK, Hultcrantz M, Schünemann HJ, et al. GRADE guidelines 33: addressing imprecision in a network meta-analysis. J Clin Epidemiol 2021;139:49−56.
[21] Schünemann HJ, Neumann I, Hultcrantz M, Brignardello-Petersen R, Zeng L, Murad MH, et al. GRADE guidance 35: update on rating imprecision for assessing contextualized certainty of evidence and making decisions. J Clin Epidemiol 2022;150:225−42.