BMJ Open Sport & Exercise Medicine

How to conduct and report checking transitivity and inconsistency in network-meta-analysis: a narrative review including practical worked examples, code and source data for sports and exercise medicine researchers

Daniel L Belavý ,[1] Svenja Kaczorowski ,[1] Tobias Saueressig ,[2] Patrick J Owen ,[3,4] Adriani Nikolakopoulou[5,6]

To cite: Belavý DL, Kaczorowski S, Saueressig T, *et al*. How to conduct and report checking transitivity and inconsistency in network-meta-analysis: a narrative review including practical worked examples, code and source data for sports and exercise medicine researchers. *BMJ Open Sport & Exercise Medicine* 2024;**10**:e002262. doi:10.1136/bmjsem-2024-002262

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjsem-2024-002262).

Accepted 25 November 2024

For numbered affiliations see end of article.

Correspondence to
Professor Daniel L Belavý; belavy@gmail.com

ABSTRACT

The use of network meta-analysis (NMA) in sport and exercise medicine (SEM) research continues to rise as it enables the comparison of multiple interventions that may not have been assessed in a single randomised controlled trial. NMA can then inform clinicians on potentially better interventions. Despite the increased use of NMA, we have observed that in the SEM field, a key challenge for author groups can be the assessment and reporting of key assumptions, in particular transitivity and consistency. This paper provides SEM researchers with a practical guide on how to approach the transitivity and consistency assumptions of NMA. Using a previously published NMA in the SEM field, we provide the statistical code, source data and worked examples to facilitate understanding and best practice of NMA in the particular field. We hope these resources result in improved conduct and reporting of NMA that ultimately leads to advances in the SEM field.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Network meta-analysis (NMA) continues to be used increasingly in sport and exercise medicine (SEM) research, a key challenge for author groups is often the assessment and reporting of the assumptions of transitivity and inconsistency.

WHAT THIS STUDY ADDS

⇒ This paper provides SEM researchers with a practical statistical code, source data and worked examples to guide their approach to transitivity and consistency assumptions.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ We hope these resources result in improved conduct and reporting of NMA that will ultimately lead to advances in the SEM field.

INTRODUCTION

The use of network meta-analysis (NMA) in sport and exercise medicine (SEM) research continues to rise (figure 1) as the field strives to produce the highest level of medical evidence.[1] NMA is a statistical technique that allows the simultaneous comparison between three or more interventions (eg, exercise training vs manual therapy vs no-intervention control) to estimate the comparative effects and can help determine best practice standards.[2] NMA can assist in choosing between alternative interventions with a thorough understanding of relative benefits and potential harms by offering one review that covers all pertinent interventions.[2] One of the main features of NMA is that it combines direct and indirect evidence from multiple studies and therefore overcomes cardinal limitations associated with pairwise meta-analysis. For example, NMA could enable an indirect comparison of exercise training and manual therapy where there were only studies available that compared either intervention to true control. Subsequently, NMA can produce estimates that are more precise than direct or indirect comparisons alone.[2]

While reporting of NMA appears to be improving in general,[3] we have observed that in the SEM field, assessment and reporting of key assumptions, in particular transitivity and consistency, can represent a challenge for author groups. A thorough assessment of these assumptions is crucial in the conduct of NMA to ensure valid effect estimates from direct and indirect evidence. Here, we provide support for SEM researchers conducting NMA to help improve the reporting of assumptions in frequentist and Bayesian NMA.

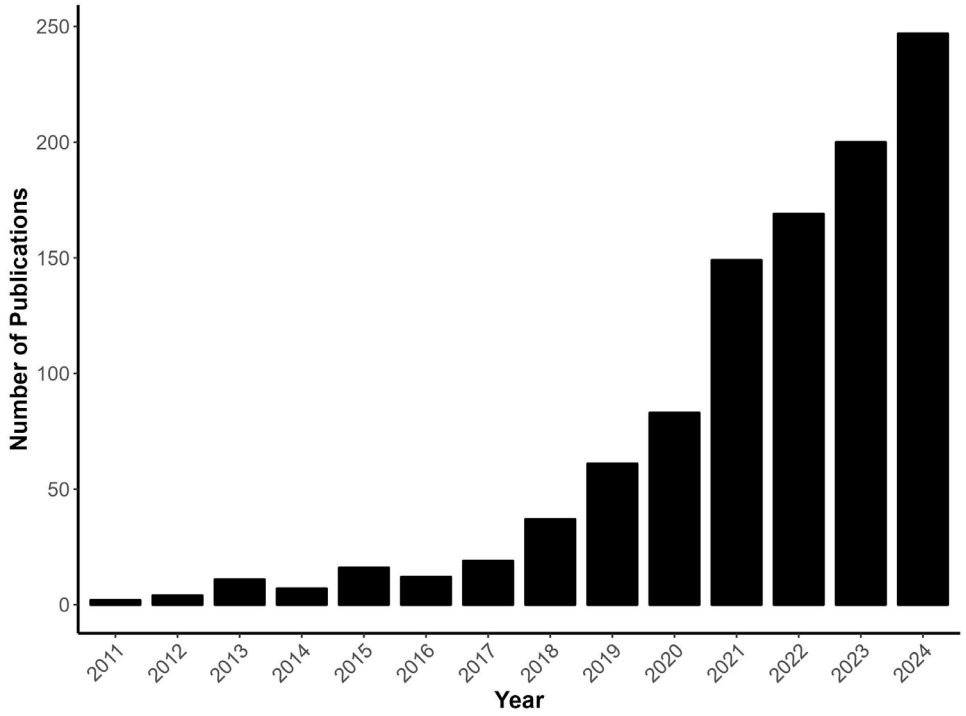

Figure 1 Recent uptake of network meta-analysis in sport and exercise medicine research. Number of sport and exercise-related publications in each year that mention network meta-analysis in the title or abstract. PubMed search: network meta-analysis[title/abstract] AND (sport OR exercise).

This paper focuses specifically on assumption checking in NMA, thus the reader should be familiar with the general concept of NMA. For a general overview and introduction to NMA, we refer readers to existing literature that is more readily understandable to non-statisticians.[4–8] We assume, and recommend, that SEM researchers will already have a (bio)statistical advisor who is experienced in NMA. Producing key outputs (eg, league tables, forest plots, box plots) is provided in our example code, yet the focus of our paper is on assumption checking and we therefore refer readers to existing guidance for these wider aspects of reporting.[6–8] For high-quality conduct of systematic reviews, we refer the reader to the Cochrane Handbook.[9]

The example data set

Here, we use a dataset from an NMA in the SEM field that examined the effects of different exercise training types in adults with chronic non-specific low back pain.[10] In this dataset, the SEM researchers examined 10 different exercise training types and collected data on subjective pain intensity, physical function (ie, disability) and mental health. For brevity, our examples herein use pain intensity data only, which is the most frequently reported outcome. To ensure readers can replicate our analyses and adapt them using their own data, we also provide the statistical code for the packages 'netmeta',[11] 'metafor (frequentist NMA meta-regression)'[12] and 'multinma'[13] for frequentist and Bayesian NMA in the R statistical environment (Version 4.32). The tabulated input data and code are openly available online (https://osf.io/k9jfa/?view_only=845cbc7df9264faa83958cd4a8f7381d) and the network plot is presented in figure 2.

Assessing the assumption of transitivity

As with pairwise meta-analysis, one key assumption of NMA is that the included studies, populations and

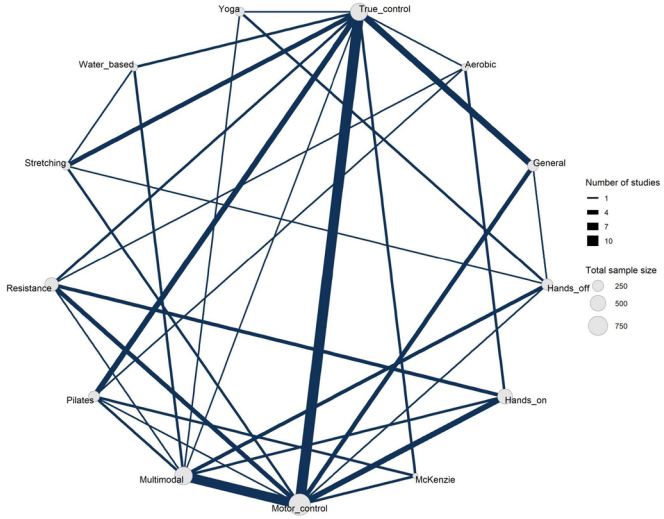

Figure 2 Network plot for our example data set. For each exercise type, node size is proportional to the number of participants in the studies with those exercise types. The thickness of the edges (lines) is proportional to the number of studies that provide data for that head-to-head comparison of exercise types. Data are drawn from a prior publication.[10] The data and statistical code are available freely online http://doi.org/10.17605/OSF.IO/K9JFA.

Table 1 Inclusion and exclusion criteria used in the prior network meta-analysis[10]

	Inclusion	Exclusion
General requirements	Published in peer-reviewed journal in any language	Grey literature
Population	Adults (≥18 years) with non-specific chronic (≥12 weeks) low back pain (localised below the costal margin and above the inferior gluteal folds, with or without leg pain)	Pain due to or associated with pregnancy, infection, tumour, osteoporosis, fracture, structural deformity (eg, scoliosis), inflammatory disorder, radicular syndrome or cauda equine syndrome
		Studies that solely recruited patients pre- or post-surgery
Intervention	Prescription of exercise training alone, without the addition of other treatments (eg, massage, ultrasound or hot and cold therapy) for at least 4 weeks of duration	
Control	Non-exercise training intervention (including true control) or another exercise training intervention	
Outcome	At least one of the outcome measures of interest: subjective pain intensity (eg, visual analogue scale), subjective physical function (eg, Oswestry Disability Index), objective trunk muscle strength (eg, lumbar extension one-repetition maximum), objective trunk muscle endurance (eg, static lumbar extension hold time), subjective analgesic pharmacotherapy use (eg, prescription medication use) or subjective mental health (eg, 36-Item Short Form Health Survey)	
Study Design	Parallel arm (individual- or cluster-designed) randomised controlled trials	Total sample size of <20 patients

interventions are 'sufficiently similar' to each other[6 7 14 15] and that the treatment effects being compared are not affected by differences in effect modifiers (variables that change the relative treatment effect).[16 17] In the planning phase of an NMA, SEM researchers should consider minimising the clinical and methodological heterogeneity of the included studies by formulating strict selection criteria regarding the population and condition, as well as the interventions under evaluation for the NMA. An example of this is shown in table 1.

On top of similarity within treatment comparisons, transitivity in NMA requires that effect modifiers are similar across comparisons. Once the data for an NMA are available, SEM researchers can create tables for the transitivity assessment of patient and trial characteristics in each study (online supplemental S1), of effect modifiers by treatment nodes (online supplemental S2) and of effect modifiers by pairwise comparisons (online supplemental S3). In general, the choice of effect modifiers should be based on existing data regarding effect modifiers that may influence the outcome of an intervention.[6 7] On this basis, SEM researchers can:

1. Judge the requirement of joint randomisability. In particular, could a mega-trial be conducted that would include all the interventions that were specified as eligible in the NMA? In our example, the interventions that are included do not differ with respect to their indication; such a trial is hypothetically possible, and thus transitivity is likely to hold. In this example, a treatment such as spinal cord stimulation should not be attempted before non-invasive treatments such as exercise are attempted. Thus, transitivity may not hold if spinal cord stimulation was an included intervention.
2. Examine study and patient characteristics across the studies, as well as potential effect modifiers (eg, disease severity), narratively and make a clinical judgement as to the importance of the differences between studies. In our example, the effect modifiers that we chose are baseline back pain, physical function and mental health. These effect modifiers were selected because they may have an impact on the relative effectiveness between interventions.[18–23] For example, patients with higher pain intensity levels at baseline may show a larger improvement than those with lower pain intensity levels at baseline.[24]
3. Inspect transitivity visually by creating box plots depicting effect modifiers for each treatment node and corresponding pairwise comparisons (online supplemental S4-5). In our example, inspection of the boxplots for baseline values of pain intensity, physical function and mental health for treatment nodes and pairwise comparisons mostly suggest that the transitivity assumption holds (for an example see figure 3).
4. If enough data are available, these analyses can be augmented by conducting tests for differences between treatment nodes or pairwise comparisons against baseline values. Analysis of variance or Kruskal-Wallis tests could be used, respectively, based on the distribution of the effect modifiers being normal or not. However, these tests may have a low statistical power and thus should not be used as a single criterion to rate transitivity. In our example, the statistical tests did not provide evidence for a difference between nodes (online supplemental S6-7); thus, based on this test and the results of all other transitivity checks, we conclude in our example that the assumption of transitivity holds.

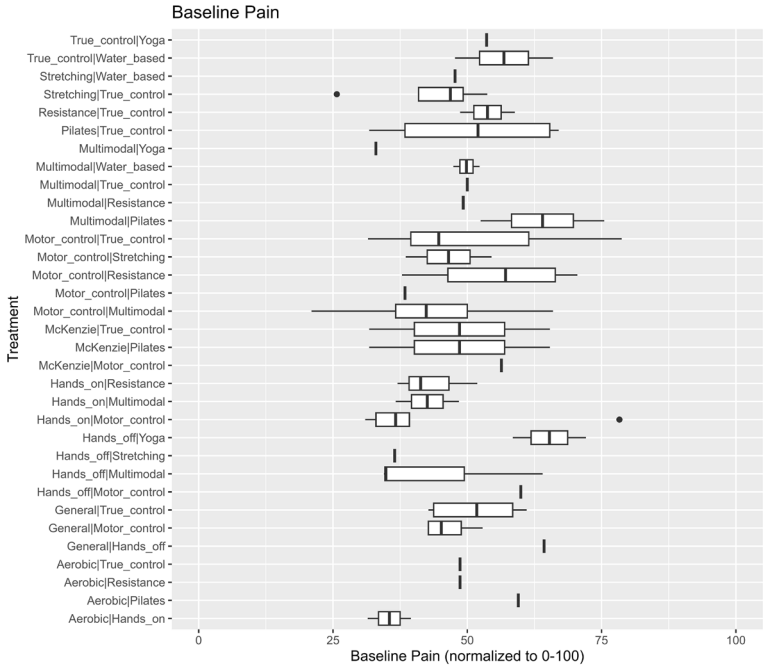

Figure 3 The figure shows the variable mean baseline pain intensity normalised to 0–100 scale on the abscissa and the treatment comparisons found in the network on the ordinate. The boxplot illustrates the data distribution by showing the median, quartiles and possible outliers. The line inside the box displays the median, and the box itself represents the middle 50% of the data (from the first to the third quartile). Any values outside of the whiskers are regarded as outliers. The 'whiskers' span the smallest and largest values within 1.5 times the IQR. All trials provided data. We assessed the similarity of the mean baseline pain intensity value of the treatment comparisons by visual inspection of the overlap of the 25th to 75th percentiles, which are largely overlapping indicating that the transitivity assumption is fulfilled. There were two outliers determined by lying beyond the whiskers of the boxplot. The comparison of Stretching versus True Control with 25.7 points and the comparison of Hands on versus Motor control with 78.3 points. These correspond to the studies: Kofotolis 2008: Hands on versus Motor control with a pooled pain intensity value of 78.31 points, and Segal-Snir 2016: True control versus Stretching with a pooled pain intensity value of 25.71 points. We can confirm the normality assumption of the analysis of variance (ANOVA) residuals by conducting the Shapiro-Wilk normality test (p=0.1981) and visually inspecting the histogram and quantile-quantile (QQ) plots of the ANOVA residuals. The p value for the F-test in the ANOVA is 0.94 and shows that there are no significant differences between nodes.

Global and local inconsistency

Consistency is the statistical extension of transitivity[6–8] and can detect potential problems with the network estimates even if the transitivity assumption holds. When we check NMA results for inconsistency, we want to know whether the effects obtained from indirect comparisons are in agreement with those from direct comparisons in the network.[6] Inconsistency methods can be subdivided into global and local, where global inconsistency is referring to inconsistency within the whole network and local inconsistency to specific loops or comparisons within the network.[6] The recommended approach in both frequentist and Bayesian NMA is to assess both global inconsistency and local inconsistency. Even if, in an analysis, there is no global inconsistency, it is still considered good practice to examine local inconsistency.[9]

Assessing the assumption of global and local inconsistency

In a frequentist NMA approach, global inconsistency can be assessed by:

- Decomposition of the Q-statistic (see online supplemental S8). In our example, we estimate a p value of 0.002, which shows statistically significant inconsistency at an α of 0.05 (online supplemental S8).

To then assess local inconsistency in a frequentist NMA:

- An analysis of side-splitting is conducted (online supplemental S9). In this approach, the direct and indirect evidence of each comparison in the network is compared with each other. Here, SEM researchers would examine the p values and CIs for discrepancy between direct and indirect evidence. In our example, four comparisons show evidence for local inconsistency.
- Inspect local inconsistency visually by creating side-split plots (figure 4 and online supplemental S10). Completely overlapping direct and indirect effect estimates and corresponding CIs indicate no local consistency. Since some of the example plots of corresponding comparisons from the side-splitting are not overlapping, the side-split plots support results from the quantitative analysis in step 1. In our particular example, the differences between direct and indirect estimates presented in figure 4 may be due to the

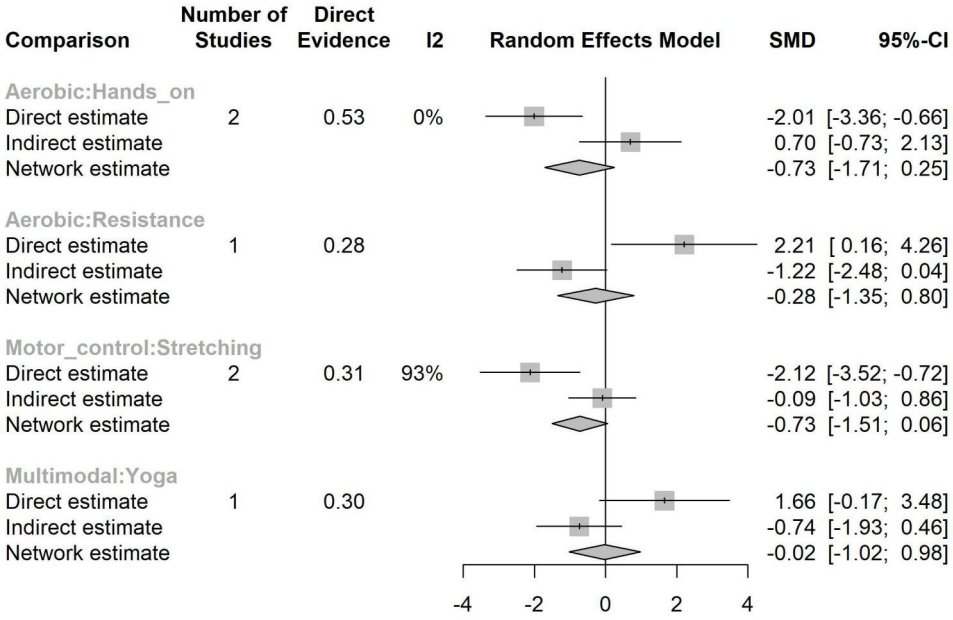

Figure 4 Frequentist tests for local inconsistency (side-splitting). Since some of the plots of the corresponding comparisons from the side-splitting are not overlapping, the side-split plots confirm our results from the numerical analysis. Please note that only the comparisons displaying a significant p value for the test for disagreement between direct and indirect evidence are presented. This decision was made to ensure the clarity and readability of the plot. The full side-splitting plot is presented in online supplemental S10. SMD, standardised mean difference.

kinds of interventions performed in RCTs of multimodal exercise, stretching exercise, aerobic exercise and the 'hands on' comparator that were finally included in the review.

In the frequentist model of our example dataset, both global (decomposition of the Q-statistic) and local methods (side-splitting analyses) demonstrated statistically significant discrepancies between direct and indirect evidence.

For global inconsistency in Bayesian NMA, the unrelated mean effects model (UME) is used. In a UME model, each relative effect is estimated separately based on direct data without the assumption of consistency.[25] The UME model is then compared with the normal (consistency) Bayesian NMA model. SEM researchers should assess:

- Whether the deviance information criterion (DIC) of the UME model is at least three or five points less than that of the Bayesian NMA model. We prefer a stricter threshold and use a three-point difference as a criterion for model comparison. In our example, the DIC differs by 0.87 and is below the threshold for inconsistency[25] (online supplemental S11).
- Residual deviance: assess whether the total residual deviance of the UME model is at least three or five points less than that of the Bayesian NMA model. In our example, the comparison between the UME and the 'normal' NMA model suggests no inconsistency based on the difference in residual deviances of 2.27 (online supplemental S11).
- Graphically: to compare the UME with the normal NMA model via deviance contribution plots (figure 5). These plots show the difference between the observed and expected deviance for each data point in the network, where the expected deviance is based on the normal NMA model (consistency model). In our example, the comparison between the UME and the 'normal' NMA model suggests inconsistency based on multiple points that fall below the line of equality in the deviance contribution plot (figure 5).

We can conclude that there is evidence for inconsistency in the Bayesian analysis and an assessment of local inconsistency is warranted to further explore this issue. For this, SEM researchers should:

- Conduct node-splitting analyses for Bayesian NMA (online supplemental S12) and evaluate inconsistency by comparing the direct and indirect estimates using a Z-test and the corresponding (Bayesian) p value. In our example, Bayesian node splitting showed statistically significant Bayesian p values for tests for disagreement between direct and indirect evidence for the following comparisons: 'aerobic' versus 'true control' (p=0.007), 'hands on' versus 'aerobic' (p=0.012), 'resistance' versus 'aerobic' (p=0.041), 'stretching' versus 'motor control' (p=0.016) and 'yoga' versus 'multimodal' (p=0.034) at an α of 0.05 (online supplemental S12).
- Similar to frequentist NMA, inspect local inconsistency visually by creating node-splitting plots (figure 6 and online supplemental S13). Bayesian node-splitting plots show the distributions of each estimate (direct, indirect, NMA) and should completely overlap in case of zero local inconsistency (figure 6 and online supplemental S13). Similar to the frequentist analysis,

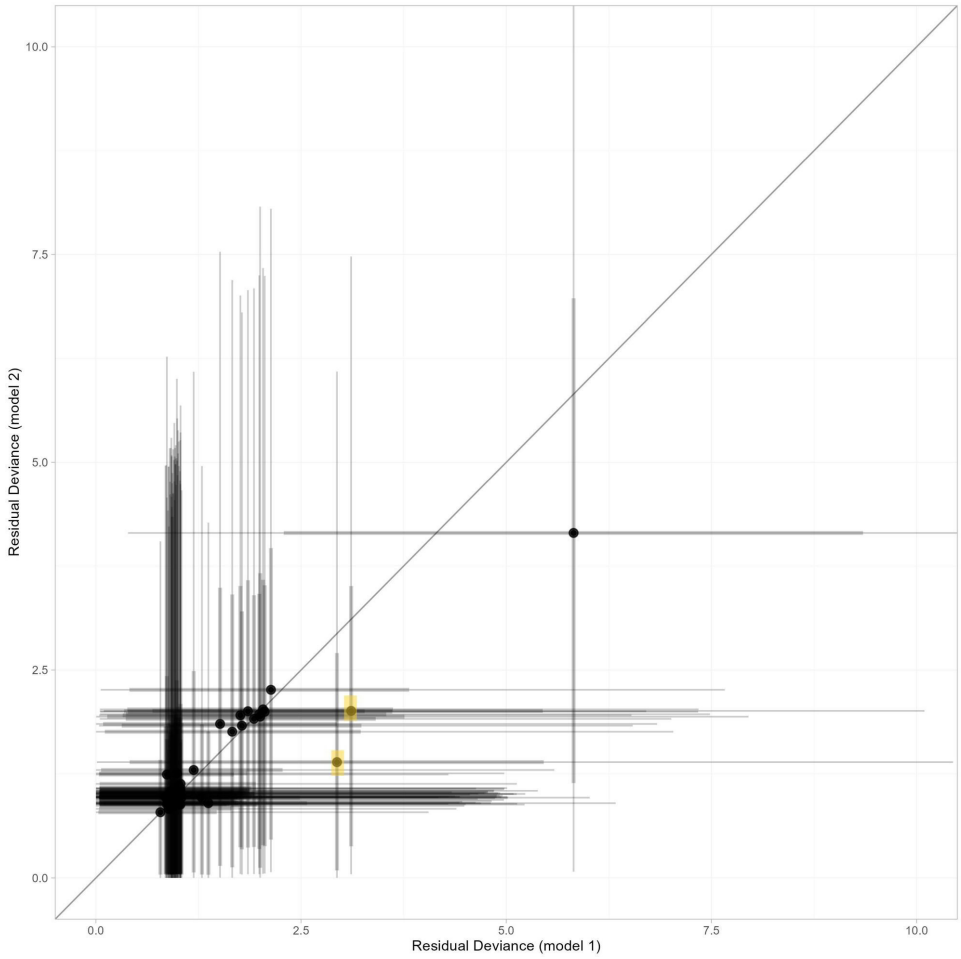

Figure 5 Deviance contribution plot. These plots can be used to assess, in Bayesian NMA whether inconsistency is present. If the points are close to the line of equality, it means the UME and the NMA models are consistent. If the points are below the line of equality, it means the UME model fits the data better than the NMA model, which indicates inconsistency.[25] Note that the NMA model is model 1 and the UME is model 2. Multiple points fall below the line of equality. That is an indication for global inconsistency. See, for example, the points marked in yellow below. Here, the inconsistency model better predicted the data points, which can be seen by an improvement in residual deviance of around 2.94–1.40=1.54 and 3.11–2.01=1.10. Improvements in residual deviance of at least 0.5 should be flagged as well. NMA, network meta-analysis; UME, unrelated mean effects.

some node splits are not overlapping, and confirm our results.

Overall, our example highlights potential inconsistencies in the NMA results and indicates that further exploration of the data may be necessary.

Potential approaches to inconsistency when it is present

For NMA, exploring inconsistency is crucial to ensure the reliability and credibility of synthesised evidence. This involves investigating potential sources of inconsistency between direct and indirect treatment comparisons. In this context, several approaches can be employed to systematically understand and address inconsistency when it arises.

Data accuracy checks: ensuring the accuracy of data extraction is paramount. Techniques such as side- or node-splitting results, along with Bayesian NMA deviance contribution plots,[25] can be employed to identify potential errors in data extraction (eg, confusing SD with the SE). These approaches help pinpoint studies that might be contributing disproportionately to inconsistency, allowing for a more targeted and informed resolution. For our example, we checked the data multiple times to ensure its correctness.

Network meta-regression models: meta-regression models applied to the network setting offer an avenue to investigate potential effect modifiers systematically. By exploring how various potential effect modifiers impact treatment effects across studies, SEM researchers can gain a deeper understanding of sources of inconsistency and refine their interpretations accordingly. In our example, we performed Bayesian network meta-regression with pooled baseline pain intensity as the continuous covariate. An independent/unrelated interaction term was used[26] and baseline pain intensity was mean-centred to allow for better interpretability. Bayesian network meta-regression showed only a significant beta coefficient (β: 0.12; 95%

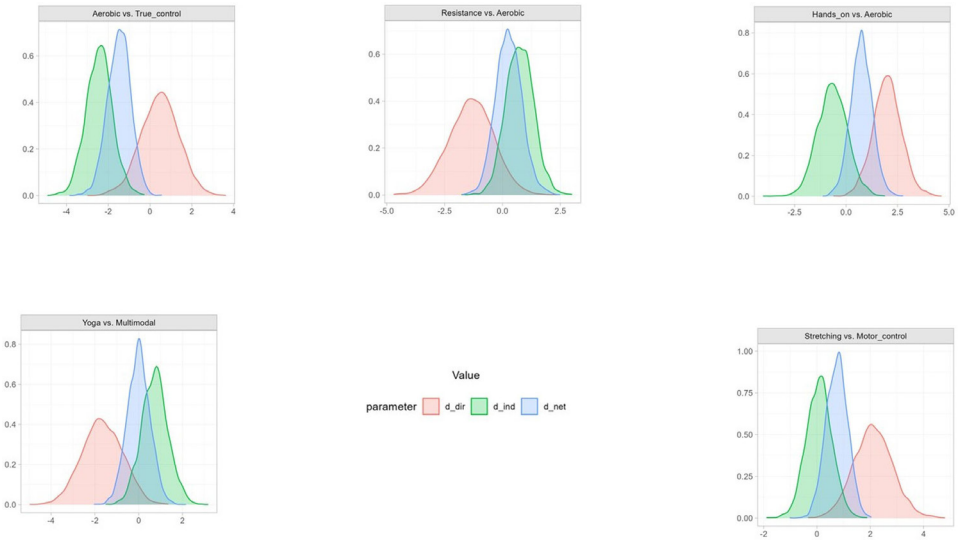

Figure 6 Bayesian tests for local inconsistency (node splitting). Please note that only the comparisons displaying a significant p value for the test for disagreement between direct and indirect evidence are presented. This decision was made to ensure the clarity and readability of the plot. In the plot, direct effects are represented in red, indirect effects in green and network meta-analysis (NMA) effects in blue. The ideal scenario occurs when all distributions overlap, indicating consistency among estimates. The full side-splitting plot is presented in online supplemental S13.

credible interval: 0.02, 0.22) for the aerobic node, indicating that baseline pain intensity explained part of the inconsistency. The interpretation is that, at least for aerobic exercise, a higher baseline pain level reduces the effectiveness of the intervention. Overall, model comparison between the base-case NMA model and the network meta-regression model did not show an improvement in model fit for Bayesian analyses (online supplemental S14). This suggests that the covariate baseline pain intensity does not explain inconsistency fully for the Bayesian analysis.

Sensitivity analyses: conducting sensitivity analyses, such as excluding studies identified as potential sources of inconsistency, provides an opportunity to assess the robustness of NMA results. This approach helps SEM researchers gauge the extent to which the synthesised evidence is dependent on specific studies and enhances the overall reliability of the findings. As an example, we excluded studies that had a total sample size of <50 participants. We only performed frequentist NMA for this analysis and found that the test for global inconsistency was no longer statistically significant (p=0.1152). Assessment of local inconsistency showed that one comparison (Pilates vs True Control) did show inconsistency (p=0.04) (online supplemental S15). This could suggest that the inclusion of studies with smaller sample sizes may have influenced the initial findings, potentially introducing heterogeneity, inconsistency and/or bias. Another explanation for the result could be that the removal of the studies with a lower sample size reduced the power of the test to detect inconsistency.

Qualitative assessment: to qualitatively address inconsistency, the Grading of Recommendations Assessment, Development and Evaluation (GRADE) framework offers potentially valuable insights.[27] Two notable GRADE-based approaches, namely, the Confidence in Network Meta-Analysis (CINeMA)[28] and the method proposed by Puhan *et al*,[29 30] provide systematic tools for assessing and interpreting the quality of evidence, allowing researchers to gain a nuanced understanding of the reliability of NMA results. It has to be noted that the CINeMA approach is currently only established for frequentist NMA. In our example, the evidence of the network would be downgraded to two levels for inconsistency because global inconsistency appears to be present.

Finally, even if all inconsistency cannot be explained, as may be the case in our example, per the PRISMA NMA[8] reporting guideline, it is considered good practice to present the analyses in full to enable members of the broader SEM field to make their own assessments.

CONCLUSION

This paper provides SEM researchers with a practical guide on how to approach the transitivity and (in)consistency assumptions of NMA. Using a previously published NMA in the SEM field, we provide the practical tools, source data and worked examples to further facilitate understanding. We hope these resources result in improved conduct and reporting of NMA that ultimately leads to advances in the SEM field.

Author affiliations
[1]Department für Pflege-, Hebammen- und Therapiewissenschaften, Hochschule für Gesundheit, Bochum, Germany
[2]Physio Meets Science GmbH, Heidelberg, Germany
[3]Eastern Health, Melbourne, Victoria, Australia
[4]Monash University, Melbourne, Victoria, Australia
[5]School of Medicine, Aristotle University of Thessaloniki, Thessalonike, Kentrikḗ Makedonía, Greece
[6]Institute of Medical Biometry and Statistics, Albert-Ludwigs-Universität Freiburg, Freiburg im Breisgau, Baden-Württemberg, Germany

X Daniel L Belavý @belavyprof and Patrick J Owen @PatrickOwenPhD

Acknowledgements We gratefully acknowledge Nana-Adjoa Kwarteng for her valuable contributions to data analysis and coding.

Contributors Guarantor: DLB. Contributions: Conceptualisation: DLB and AN. Methodology: DLB, TS, PJO and AN. Software: TS. Validation: SK Formal analysis: TS. Investigation: all. Data curation: PJO, DLB and TS. Writing—original draft: DLB, TS and SK. Writing—review and editing: all authors. Visualisation: TS. Funding acquisition: DLB. Guarantor: DLB.

Funding This work was supported in part by grant 01KG2313 (to DLB; registration https://osf.io/39zfv/) from the German Federal Ministry of Education and Research (BMBF). AN was supported by the Emmy Noether Programme of the German Research Foundation (grant number NI 2226/1-1).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval Not applicable.

Provenance and peer review Not commissioned; externally peer reviewed.

ORCID iDs
Daniel L Belavý http://orcid.org/0000-0002-9307-832X
Svenja Kaczorowski http://orcid.org/0000-0001-8794-3252
Tobias Saueressig http://orcid.org/0000-0002-3332-3856
Patrick J Owen http://orcid.org/0000-0003-3924-9375

REFERENCES

1 Faltinsen EG, Storebø OJ, Jakobsen JC, *et al*. Network meta-analysis: the highest level of medical evidence? *BMJ Evid Based Med* 2018;23:56–9.
2 Higgins J, Thomas J, Chandler J, eds. *Cochrane handbook for systematic reviews of interventions version 6.0*. Cochrane, 2019.
3 Veroniki AA, Tsokani S, Zevgiti S, *et al*. Do reporting guidelines have an impact? Empirical assessment of changes in reporting before and after the PRISMA extension statement for network meta-analysis. *Syst Rev* 2021;10:246.
4 Catalá-López F, Tobías A, Cameron C, *et al*. Network meta-analysis for comparing treatment effects of multiple interventions: an introduction. *Rheumatol Int* 2014;34:1489–96.
5 Dias S, Caldwell DM. Network meta-analysis explained. *Arch Dis Child Fetal Neonatal Ed* 2019;104:F8–12.
6 Rouse B, Chaimani A, Li T. Network meta-analysis: an introduction for clinicians. *Intern Emerg Med* 2017;12:103–11.
7 Chaimani A, Salanti G, Leucht S, *et al*. Common pitfalls and mistakes in the set-up, analysis and interpretation of results in network meta-analysis: what clinicians should look for in a published article. *Evid Based Ment Health* 2017;20:88–94.
8 Hutton B, Salanti G, Caldwell DM, *et al*. The PRISMA Extension Statement for Reporting of Systematic Reviews Incorporating Network Meta-analyses of Health Care Interventions: Checklist and Explanations. *Ann Intern Med* 2015;162:777–84.
9 Higgins J, Thomas J, Chandler J, eds. *Cochrane handbook for systematic reviews of interventions*. Cochrane, 2021.
10 Owen PJ, Miller CT, Mundell NL, *et al*. Which specific modes of exercise training are most effective for treating low back pain? Network meta-analysis. *Br J Sports Med* 2020;54:1279–87.
11 Balduzzi S, Rücker G, Nikolakopoulou A, *et al*. netmeta: An R Package for Network Meta-Analysis Using Frequentist Methods. *J Stat Softw* 2023;106:1–40.
12 Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *J Stat Softw* 2010;36.
13 Phillippo D. Dmphillippo/multinma: v0.3.0. 2021.
14 Jansen JP, Naci H. Is network meta-analysis as valid as standard pairwise meta-analysis? It all depends on the distribution of effect modifiers. *BMC Med* 2013;11:159.
15 Ades AE, Welton NJ, Dias S, *et al*. Twenty years of network meta-analysis: Continuing controversies and recent developments. *Res Synth Methods* 2024;15:702–27.
16 Donegan S, Williamson P, D'Alessandro U, *et al*. Assessing key assumptions of network meta-analysis: a review of methods: Assessing assumptions of network meta-analysis. *Res Synth Methods* 2013;4:291–323.
17 Daly C, Dias S, Welton NJ, *et al*. Meta-analysis: guideline methodology document 1. In: *NICE Guidelines Technical Support Unit*. 2021: 39.
18 Boonstra AM, Reneman MF, Waaksma BR, *et al*. Predictors of multidisciplinary treatment outcome in patients with chronic musculoskeletal pain. *Disabil Rehabil* 2015;37:1242–50.
19 Cecchi F, Pasquini G, Paperini A, *et al*. Predictors of response to exercise therapy for chronic low back pain: result of a prospective study with one year follow-up. *Eur J Phys Rehabil Med* 2014;50:143–51.
20 Steffens D, Hancock MJ, Maher CG, *et al*. Prognosis of chronic low back pain in patients presenting to a private community-based group exercise program. *Eur Spine J* 2014;23:113–9.
21 van der Hulst M, Vollenbroek-Hutten MMR, Ijzerman MJ. A systematic review of sociodemographic, physical, and psychological predictors of multidisciplinary rehabilitation-or, back school treatment outcome in patients with chronic low back pain. *Spine (Phila Pa 1976)* 2005;30:813–25.
22 Adnan R, Van Oosterwijck J, Cagnie B, *et al*. Determining Predictive Outcome Factors for a Multimodal Treatment Program in Low Back Pain Patients: A Retrospective Cohort Study. *J Manipul Physiol Ther* 2017;40:659–67.
23 Dhondt E, Van Oosterwijck J, Cagnie B, *et al*. Predicting treatment adherence and outcome to outpatient multimodal rehabilitation in chronic low back pain. *J Back Musculoskelet Rehabil* 2020;33:277–93.
24 Konrad KL, Baeyens J-P, Birkenmaier C, *et al*. The effects of whole-body electromyostimulation (WB-EMS) in comparison to a multimodal treatment concept in patients with non-specific chronic back pain-A prospective clinical intervention study. *PLoS ONE* 2020;15:e0236780.
25 Daly C, Downing BC, Welton NJ. A practical guide to inconsistency checks in bayesian network meta-analysis. 2022.
26 Efthimiou O, Debray TPA, van Valkenhoef G, *et al*. GetReal in network meta-analysis: a review of the methodology. *Res Synth Methods* 2016;7:236–63.
27 Guyatt G, Oxman AD, Akl EA, *et al*. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. *J Clin Epidemiol* 2011;64:383–94.
28 Nikolakopoulou A, Higgins JPT, Papakonstantinou T, *et al*. CINeMA: An approach for assessing confidence in the results of a network meta-analysis. *PLoS Med* 2020;17:e1003082.
29 Puhan MA, Schünemann HJ, Murad MH, *et al*. A GRADE Working Group approach for rating the quality of treatment effect estimates from network meta-analysis. *BMJ* 2014;349.
30 Brignardello-Petersen R, Bonner A, Alexander PE, *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.