

Prevalence of Inconsistency and Its Association With Network Structural Characteristics in 201 Published Networks of Interventions

Areti Angeliki Veroniki (✉ averoniki@uoi.gr)

University of Ioannina

Sofia Tsokani

University of Ioannina

Ian R. White

University College London

Guido Schwarzer

University of Freiburg

Gerta Rücker

University of Freiburg

Dimitris Mavridis

University of Ioannina

Julian P.T. Higgins

University of Bristol

Georgia Salanti

University of Bern

Research Article

Keywords: indirect evidence, network meta-analysis, chi-squared test, between-study variance, incoherence

Posted Date: April 16th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-403088/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: Network meta-analysis (NMA) has attracted growing interest in evidence-based medicine. Consistency between different sources of evidence is fundamental to the reliability of the NMA results. The purpose of the present study was to estimate the prevalence of inconsistency and describe its association with different NMA characteristics.

Methods: We updated our collection of NMAs with articles published up to July 2018. We included networks with randomised clinical trials, at least four treatment nodes, at least one closed loop, a dichotomous primary outcome, and available arm-level data. We assessed consistency using the design-by-treatment interaction (DBT) model. We estimated the prevalence of inconsistency and its association with different network characteristics (e.g., number of studies, treatments, treatment comparisons, loops), and evaluated heterogeneity in NMA and DBT models.

Results: We included 201 published NMAs. The p-value of the design-by-treatment interaction (DBT) model was lower than 0.05 in 14% of the networks and lower than 0.10 in 20% of the networks. Networks comparing few interventions in many studies were more likely to have small DBT p-values (less than 0.10), which is probably because they yielded more precise estimates and power to detect differences between designs was higher. In the presence of inconsistency (DBT p-value lower than 0.10), the consistency model displayed higher heterogeneity than the DBT model.

Conclusions: Our findings show that inconsistency was more frequent than what would be expected by chance, suggesting that researchers should devote more resources to exploring how to mitigate inconsistency. The results of this study highlight the need to develop strategies to detect inconsistency (because of the relatively high prevalence of inconsistency in published networks), and particularly in cases where the existing tests have low power.

1. Background

Network meta-analysis (NMA) is a useful approach for exploring effects of multiple interventions by simultaneously synthesizing direct and indirect evidence, and in recent years the number of published NMAs has grown continually[1, 2]. The reliability of inferences from NMA depends on the comparability of studies evaluating the multiple interventions of interest[3–5]. NMA results are valid only when the transitivity assumption holds, i.e., the distribution of effect modifiers is similar across intervention comparisons. Lack of transitivity can create statistical disagreement between the information coming from direct and indirect sources of evidence, that is inconsistency. The notion of inconsistency also refers to the disagreement between evidence coming from different designs (comparisons of different sets of interventions across studies). Several statistical methods exist to assess inconsistency in the network as a whole or locally on specific comparisons or loops of evidence (i.e., paths in the network of interventions that start and end at the same node)[6–12]. To date, the design-by-treatment interaction (DBT) model is

the only method that both provides a global assessment of inconsistency for a network and is insensitive to the parameterization of studies with multiple arms[6, 8].

The majority of NMAs published in the medical literature in the recent years examine whether the prerequisite assumptions in NMA are met[1, 2, 13, 14]. However, a number of reviews still combine direct and indirect evidence in a network of interventions without evaluating the condition of consistency or despite evidence of inconsistency[15, 16]. Empirical findings for dichotomous outcomes suggest that inconsistency is present in one in ten loops of evidence when a network is evaluated in multiple parts separately and in one in eight networks when a network is evaluated as a whole[10]. It is encouraging though that NMAs increasingly discuss transitivity and/or inconsistency (0% in 2005 vs 86% in 2015), and use appropriate methods to test for inconsistency (14% in 2006 vs 74% in 2015)[2]. Another important consideration when conducting an NMA is that there is an inverse association between heterogeneity and statistical power to detect inconsistency[3]. The larger the heterogeneity, the less precise the direct and indirect estimates are, and hence statistical inconsistency may not be evident even when it is present. Empirical evidence using 40 networks of interventions suggested that increased heterogeneity was associated with low detection rates of inconsistency, and that the consistency model displayed higher heterogeneity than the inconsistency model[10]. Also, the same study showed that the choice of the heterogeneity estimator can influence inferences about inconsistency particularly when few studies are available. In general inconsistency tests have low power to detect inconsistency, and power may vary depending on the network's characteristics (e.g., number of patients and studies)[17].

The purpose of the present study was to estimate the percentage of NMAs for which strong evidence against the hypothesis of consistency is evident. We updated our previous empirical evaluation using a larger sample of published NMAs with dichotomous outcome data[10]. We also aimed to describe the association between evidence against consistency and the NMA structural characteristics, such as number of studies, interventions, and loops. We finally aimed to evaluate heterogeneity in consistency and inconsistency models.

2. Methods

2.1 Eligibility criteria for network database

The collection of published NMAs used in this paper has been described elsewhere[1, 2]. We included networks published in two distinct periods: 1) up to December 2015 (including NMAs identified from our previous search up to April 2015[18], and from our updated search up to December 2015), and 2) between 2017 and 2018 (from our updated search up to July 2018[19]). NMAs were eligible if they included only randomised clinical trials, had at least four intervention nodes (including placebo) in the network, had conducted any form of valid indirect comparison or NMA, included at least one loop, and had a dichotomous primary outcome with available arm-level data.

2.1 Synthesis

We performed a descriptive analysis of the eligible networks regarding the following characteristics: number of included studies, interventions, comparisons with direct evidence, presence of at least one intervention comparison informed by a single study, multi-arm studies, loops, number of unique designs, presence of complex interventions (as defined by Welton *et al.*[20]), type of outcome, and type of intervention comparisons[2, 21].

We assessed consistency in each network using the DBT model that evaluates the entire network as a whole and encompasses the potential conflict between studies including different sets of interventions, named 'designs'[6, 8]. In this model we synthesised evidence in a way that reflects the extra variability due to inconsistency (i.e., beyond what is expected by heterogeneity or random error)[22], and encompassed the potential conflict between studies with different sets of interventions[6, 8]. We assessed evidence against the hypothesis of consistency based on the p-value of the DBT test (see Appendix 1). Since the tests of inconsistency are known to have low power[17, 23], and considering that empirical evidence showed that 10% of loops are inconsistent[10], we decided to use along with the commonly used cut-off p-value of 0.05, the cut-off p-value of 0.10.

We estimated the prevalence of NMAs for which evidence or strong evidence against the hypothesis of consistency was evident (at both 0.05 and 0.10 thresholds) and explored its association with network structural characteristics. We present scatterplots and box plots for the aforementioned descriptive characteristics against the p-value of the DBT test. We visually assessed if inconsistency rate changed per year of study publication in a stacked bar plot. To explore the association between the DBT p-value and prevalence of inconsistency with estimation of heterogeneity in consistency and inconsistency models, we plotted the estimated between-study standard deviation values under the consistency and inconsistency models. We used a different colour scheme for each network to indicate strong evidence against the consistency hypothesis at 0.05 and 0.10 thresholds.

Among the several approaches that have been suggested to estimate the between-study variance we selected the popular DerSimonian and Laird (DL) method and the restricted maximum likelihood (REML) method which has been shown to be a better alternative[24, 25]. For completeness, we investigated the impact of both ways to estimate the between-study variance on the consistency evaluation. We distinguish heterogeneity in the consistency and inconsistency models as: a) representing within- and between-design heterogeneity in the consistency model, and b) representing within-design heterogeneity only in the inconsistency model. We conducted both consistency and inconsistency models in Stata and R using the *network*[26] suite of commands and *netmeta*[27] package, respectively. We used both Stata and R software, since at the time of conducting the analyses the REML estimator for heterogeneity was available in the *network*[26] command in Stata only, and the DL estimator in the *netmeta*[27] R package. Currently, both REML and DL estimators for heterogeneity are available in the *netmeta*[27] package. We also calculated the I-squared statistic for each network using the *netmeta* package in R and assessed its association with the DBT p-values in a scatterplot.

3. Results

3.1 Description of the network database

From the 456 total NMAs identified from our previous search[2], we located 105 NMAs satisfying the eligibility criteria. Using the same process for the years 2015 (April 2015 to December 2015), 2017 and 2018 we also included another 96 NMAs. Overall, we included 201 NMAs that fulfilled the eligibility criteria (Appendix Fig. 1).

The median number of studies per network was 20 (IQR 13, 35), and the median number of interventions per network was seven (IQR 5, 9). Multi-arm trials were included in 142 networks (70%), with a median number of one multi-arm study (IQR 0, 3). The median number of direct intervention comparisons in the included networks was 25 (IQR 16, 42), whereas the median number of unique designs was 13 (IQR 8, 23). Most networks included at least one comparison (186 networks, 92.5%) informed by a single study. The median number of loops across networks was three (IQR 2, 7), and the median number of inconsistency parameters per network was four (IQR 2, 7) (see also Appendix 1). The median I-squared statistic was 30% (IQR 0%, 59%).

3.2 Prevalence of inconsistency

At the 0.05 threshold, strong evidence against the consistency hypothesis was detected in 28 (14%) networks. At the 0.10 level, strong evidence against the consistency hypothesis was detected in 39 (20%) networks of the 201 total networks. Changing from REML to DL estimators for heterogeneity had only a minor impact on the prevalence of inconsistency (see Appendix Table 1)[24]. Most DBT p-values were considerably higher than 0.10, irrespective of heterogeneity estimator. No change in the prevalence of inconsistency was detected across years (Appendix Fig. 2).

In the following, results are presented at the 0.05 threshold and according to the REML estimation method for the between-study variance. Results according to the DL estimator are presented in the supplementary files.

3.2.1 Evidence of inconsistency across different network structural characteristics

Lower p-values in the DBT test were more likely in networks with many studies, many direct intervention comparisons, and many designs (Appendix Fig. 3). However, these associations were rather weak (studies in the network: correlation coefficient = -0.08 [p-value = 0.25], direct intervention comparisons in the network: correlation coefficient = -0.07 [p-value = 0.33], designs in the network: correlation coefficient = -0.03 [p-value = 0.67]) and Fig. 1 shows that most factors do not greatly influence the p-value of inconsistency. A weak association was found between the ratio of the number of studies to the number of interventions the p-values in DBT (correlation coefficient = -0.07 [p-value = 0.31]; see also Appendix Fig. 4 for results using the DL heterogeneity estimator). This was expected as power in detecting inconsistency is higher in networks with many studies per intervention comparison. It should also be

noted that networks with few studies and many interventions were associated with larger heterogeneity, which can mask detection of inconsistency (Appendix Fig. 5).

The type of outcome (p-value = 0.86), type of intervention comparisons in the network (p-value = 0.75), and the presence of complex interventions (p-value = 0.08) did not suggest important differences in the distribution of the p-values calculated in DBT. Similarly, the inclusion of at least one intervention comparison with a single study in the network did not affect the assessment of the global inconsistency in the network (p-value = 0.57) (Fig. 2 and Appendix Fig. 6).

3.3 Heterogeneity in consistency and inconsistency models

An increase in the I-squared statistic was associated with a decrease in the DBT p-value (Appendix Fig. 7). In Fig. 3, we plot the estimated between-study standard deviation values under the consistency and inconsistency models by levels of evidence against the consistency hypothesis (see also Appendix Fig. 8 for results using the DL heterogeneity estimator). Evidence against the consistency assumption was associated with heterogeneity being larger in the consistency model compared to the inconsistency model.

4. Discussion

Our findings suggest that evidence of inconsistency was at least twice as frequent as what would be expected by chance if all networks were truly consistent (when we would expect one in 20 networks for 0.05 level, and one in 10 networks for 0.10 level under the null hypothesis of no inconsistent networks in our sample). Overall, evidence against the hypothesis of consistency (as defined by the p-value of the DBT test) in NMAs with dichotomous data was evident in one in seven networks using the threshold of 0.05. Taking into consideration the low power of the inconsistency test, and in particular the DBT model that has more degrees of freedom in contrast to other inconsistency tests[16, 17], we decided to use also the threshold of 0.10, where inconsistency was prevalent in one in five networks. Considering that the observed inconsistent NMAs are at 14% of the networks, we expect that the truly inconsistent networks range between 12% and 20%, assuming the test has a perfect type I error at 0.05 and power ranges between 50% and 80%.

Our study showed that structural network characteristics only weakly impact the detection of inconsistency. In particular, we found a mild association between networks including both a high number of studies and a small number of interventions or loops, and lower p-values of the DBT test. This is probably due to the increased power to detect inconsistency in these types of networks. Another key finding of our study was that an important drop in heterogeneity when moving from the consistency to the inconsistency model is associated with evidence of inconsistency. This suggests that heterogeneity estimated in the consistency model may account for discrepancies between direct and indirect evidence in the network. Results were overall consistent among DL and REML heterogeneity estimators.

To the best of our knowledge, this is the largest empirical study used to evaluate the prevalence of inconsistency in networks of trial evidence. Overall, our findings are aligned with our previous study[10], where we evaluated 40 networks of interventions. The present research study includes five times the number of networks included in our previous review, and the exploration of multiple structural network characteristics. In this study, we found slightly higher empirical rates of inconsistency compared with our previous study (i.e., 14% of 201 networks vs. 13% of the 40 networks), suggesting that researchers should devote more resources to exploring how to mitigate inconsistency.

Our study has a few limitations worth noting. First, for the empirical assessment of consistency, we evaluated articles with dichotomous outcome data restricting to the odds ratio effect measure. We expect our findings to be generalisable to other effect measures. Although our previous empirical study showed that in some cases inconsistency was reduced when moving from one effect measure to another, overall, the detected inconsistency rates were similar for different effect measures[10]. For completeness it would be interesting to carry out an empirical study for continuous outcomes to examine possible differences in inconsistency between mean differences, standardized mean differences and ratios of means. Second, in the present study we considered a common within-network heterogeneity. This is often clinically reasonable and statistically convenient. Since most direct intervention comparisons in networks comprise only few studies, sharing the same amount of heterogeneity allows such comparisons to borrow strength from the entire network. However, assuming common within-network heterogeneity, intervention comparisons with a smaller heterogeneity than that of the remaining network will be associated with a larger reported uncertainty around their summary effect, compared to what would be accurate. In such a case, the chances of detecting inconsistency decrease. Although assuming a common within-network heterogeneity can underestimate inconsistency, it better reflects how summary effects are combined in an NMA in practice. Alternatively, when heterogeneity is believed to vary across comparisons, different heterogeneity parameters can be built into the model, but need to be restricted to conform to special relationships according to the consistency assumption[28]. Third, we assessed detection of inconsistency based on a threshold of the DBT p-value, which reflects common practice, and ignored the actual differences between the different designs and the direct and indirect estimates. However, to avoid “vote-counting” of strong evidence against the consistency hypothesis we also explored the distribution of the DBT p-values according to several network structural characteristics. Fourth, we did not exclude potential outlier networks, since this was outside of the scope of the study.

In a systematic review and NMA, investigators should interpret strong evidence against the consistency hypothesis very carefully and be aware that inconsistency in a network can be absorbed into estimates of heterogeneity. Given that the descriptive prevalence of inconsistency is frequent in published NMAs, authors should be more careful in the interpretation of their results. Confidence in the findings from NMA should always be evaluated, using for example CINeMA[29] (confidence in network meta-analysis) or GRADE (Grading of Recommendations Assessment, Development, and Evaluation) for NMA approaches[30]. Since inconsistency tests may lack power to identify true inconsistency, we recommend to avoid interpreting ‘no evidence for inconsistency’ as ‘no inconsistency’. We also recommend using both a global (e.g., the DBT model) and a local approach (e.g., loop-specific approach[10] or node-splitting[7]

method) for the assessment of inconsistency in a network, before concluding about the absence or presence of inconsistency. However, detection of inconsistency often prompts authors to choose only direct evidence, which is often perceived as less prone to bias, disregarding the indirect information[23]. It is advisable though, instead of selecting between the two sources of evidence, to try to understand and explore possible sources of inconsistency and refrain from publishing results based on inconsistent evidence[5, 31].

NMA is increasingly conducted and although assessment of the required assumptions has improved in recent years, there is room for further improvement[1, 2]. Systematic reviews and NMA protocols should present methods for the evaluation of inconsistency and define strategies to be followed when inconsistency is present. The studies involved in an NMA should also be compared with respect to the distribution of effect modifiers across intervention comparisons. Authors should follow the PRISMA (Preferred Reporting Items for Systematic Review and Meta-analysis)-NMA guidelines[32] and report their inconsistency assessment results, as well as the potential impact of inconsistency in their NMA findings.

5. Conclusion

This research provides empirical evidence on the prevalence of inconsistency. Our findings show that evidence of inconsistency is more frequent than would be expected by chance if all networks were consistent. This suggests that inconsistency should be appropriately explored. Detection of inconsistency was mildly sensitive to various network characteristics and their combination. In particular, networks with a high number of studies, and a small number of interventions had larger power to detect inconsistency. Also, inconsistency was likely to manifest as extra heterogeneity when the consistency model was fitted. Lower estimates of heterogeneity in the inconsistency model compared with the consistency model were associated with higher rates of detection of inconsistency. Overall, there was a good empirical agreement of inconsistency when different heterogeneity estimation methods were used.

Given that inconsistency is frequent in nature (with up to 20% of networks expected to be inconsistent), investigators should be more careful in the interpretation of their results. Our results highlight the need for a widespread use of tools that assess confidence in the NMA findings, such as CINeMA[29], that factor-in inconsistency concerns in the interpretation. It is also essential to develop strategies to detect inconsistency, particularly in cases where the existing tests have low power. Additional investigation is needed to evaluate the performance of other inconsistency methods (e.g., node-splitting[7], Lu and Ades method[12]) and to understand their power under different meta-analytical scenarios. Further investment is required in developing methods to potentially deal with inconsistent networks (e.g., network meta-regression approaches, NMA methods for classes of interventions) and to educate researchers about their use.

List Of Abbreviations

DBT: Design-By-Treatment

DL: DerSimonian and Laird

IQR: Interquartile Range

NMA: Network meta-analysis

REML Restricted Maximum Likelihood

Declarations

ACKNOWLEDGEMENTS

We thank Alissa Epworth for drafting our search strategies, de-duplicating search results and obtaining full-text articles. We thank Patricia Rios for helping organize the identified articles in the Synthesi.SR program (<http://www.breakthroughkt.ca/login.php>). We also thank Eirini Pagkalidou for helping screen studies for inclusion. Finally, we thank Stella Zevgiti for helping screen studies for inclusion and abstract binary data from included network meta-analyses.

DECLARATION OF CONFLICTING INTERESTS

The authors declare that they have no competing interests.

AVAILABILITY OF DATA AND MATERIALS

The datasets used and/or analysed during the current study are available through the *nmadb* R package (Network Meta-Analysis Database API; <https://cran.r-project.org/web/packages/nmadb/index.html>)

AUTHORS' CONTRIBUTIONS

AAV conceptualized the study, designed and coordinated the study, screened literature for inclusion, participated in all the calibration exercises of data collection, helped resolve data discrepancies, analysed the data, interpreted the findings and wrote the manuscript. ST screened the literature search results, abstracted, coded, cleaned and analysed the data, interpreted the findings, and helped to write the draft manuscript. IRW, GSc, GR, DM and JPTH helped conceptualize the study. GSa conceptualized and designed the study. All authors read, provided substantial edits and/or comments on the manuscript, and approved the final version of the manuscript that was submitted for publication.

FUNDING

This study was funded by a Canadian Institutes of Health Research (CIHR) Catalyst Grant (Grant # 342169). The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; or decision to submit the manuscript for publication.

AAV, ST, and DM were funded from the European Union's Horizon 2020 [No. 754936]. GSa has been supported by SNSF grant agreement 320030_179158. GR was funded by the German Research Foundation (DFG), grant number RU 1747/1-2. IRW was supported by the Medical Research Council Programme MC_UU_12023/21.

ETHICAL DECLARATION

Not applicable.

References

1. Petropoulou M, Nikolakopoulou A, Veroniki AA, Rios P, Vafaei A, Zarin W, Giannatsi M, Sullivan S, Tricco AC, Chaimani A *et al*: **Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015.** *Journal of clinical epidemiology* 2017, **82**:20-28.
2. Zarin W, Veroniki AA, Nincic V, Vafaei A, Reynen E, Motiwala SS, Antony J, Sullivan SM, Rios P, Daly C *et al*: **Characteristics and knowledge synthesis approach for 456 network meta-analyses: a scoping review.** *BMC medicine* 2017, **15**(1):3.
3. Salanti G: **Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool.** *Res Synth Methods* 2012, **3**(2):80-97.
4. Caldwell DM, Ades AE, Higgins JPT: **Simultaneous comparison of multiple treatments: combining direct and indirect evidence.** *BMJ* 2005, **331**(7521):897-900.
5. 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.
6. White IR, Barrett JK, Jackson D, Higgins JPT: **Consistency and inconsistency in network meta-analysis: model estimation using multivariate meta-regression.** *Res Synth Methods* 2012, **3**(2):111-125.
7. Dias S, Welton NJ, Caldwell DM, Ades AE: **Checking consistency in mixed treatment comparison meta-analysis.** *Stat Med* 2010, **29**(7-8):932-944.
8. Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR: **Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies.** *Res Synth Methods* 2012, **3**(2):98-110.
9. Bucher HC, Guyatt GH, Griffith LE, Walter SD: **The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials.** *J Clin Epidemiol* 1997, **50**(6):683-691.
10. Veroniki AA, Vasiliadis HS, Higgins JPT, Salanti G: **Evaluation of inconsistency in networks of interventions.** *Int J Epidemiol* 2013, **42**(1):332-345.
11. Caldwell DM, Welton NJ, Ades AE: **Mixed treatment comparison analysis provides internally coherent treatment effect estimates based on overviews of reviews and can reveal inconsistency.** *J Clin*

- Epidemiol* 2010, **63**(8):875-882.
12. Lu G, Ades AE: **Assessing Evidence Inconsistency in Mixed Treatment Comparisons.** *JASA* 2006, **101**(474):447-459.
 13. Song F, Xiong T, Parekh-Bhurke S, Loke YK, Sutton AJ, Eastwood AJ, Holland R, Chen YF, Glenny AM, Deeks JJ *et al.*: **Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study.** *BMJ* 2011, **343**:d4909.
 14. Bafeta A, Trinquart L, Seror R, Ravaud P: **Analysis of the systematic reviews process in reports of network meta-analyses: methodological systematic review.** *BMJ* 2013, **347**:f3675.
 15. Donegan S, Williamson P, Gamble C, Tudur-Smith C: **Indirect comparisons: a review of reporting and methodological quality.** *PLoS One* 2010, **5**(11):e11054.
 16. Song F, Clark A, Bachmann MO, Maas J: **Simulation evaluation of statistical properties of methods for indirect and mixed treatment comparisons.** *BMC Med Res Methodol* 2012, **12**:138.
 17. Veroniki AA, Mavridis D, Higgins JPT, Salanti G: **Characteristics of a loop of evidence that affect detection and estimation of inconsistency: a simulation study.** *BMC Med Res Methodol* 2014, **14**:106.
 18. Papakonstantinou T: **nmadb: Network Meta-Analysis Database API.** <https://CRAN.R-project.org/package=nmadb> 2019.
 19. Veroniki A.A., Tsokani S., Zevgiti S., Pagkalidou I., Kontouli K.M., Ambarcioglu P., Pandis N., Lunny C., Nikolakopoulou A., Papakonstantinou T. *et al.*: *Systematic Reviews* 2021.
 20. Welton NJ, Caldwell DM, Adamopoulos E, Vedhara K: **Mixed treatment comparison meta-analysis of complex interventions: psychological interventions in coronary heart disease.** *Am J Epidemiol* 2009, **169**(9):1158-1165.
 21. Turner RM, Jackson D, Wei Y, Thompson SG, Higgins JPT: **Predictive distributions for between-study heterogeneity and simple methods for their application in Bayesian meta-analysis.** *Stat Med* 2015, **34**(6):984-998.
 22. Dias S, Welton NJ, Sutton AJ, Caldwell DM, Lu G, Ades AE: **Evidence synthesis for decision making 4: inconsistency in networks of evidence based on randomized controlled trials.** *Med Decis Making* 2013, **33**(5):641-656.
 23. Song F, Harvey I, Lilford R: **Adjusted indirect comparison may be less biased than direct comparison for evaluating new pharmaceutical interventions.** *J Clin Epidemiol* 2008, **61**(5):455-463.
 24. Veroniki AA, Jackson D, Viechtbauer W, Bender R, Bowden J, Knapp G, Kuss O, Higgins JPT, Langan D, Salanti G: **Methods to estimate the between-study variance and its uncertainty in meta-analysis.** *Res Synth Methods* 2016, **7**(1):55-79.
 25. Langan D, Higgins JPT, Jackson D, Bowden J, Veroniki AA, Kontopantelis E, Viechtbauer W, Simmonds M: **A comparison of heterogeneity variance estimators in simulated random-effects meta-analyses.** *Research synthesis methods* 2018.
 26. White IR: **Network meta-analysis.** *Stata Journal* 2015, **15**:1-34.

27. **netmeta: Network Meta-Analysis using Frequentist Methods**
28. Lu G, Ades A: **Modeling between-trial variance structure in mixed treatment comparisons.** *Biostatistics* 2009, **10**(4):792-805.
29. Nikolakopoulou A, Higgins JPT: **CINeMA: An approach for assessing confidence in the results of a network meta-analysis.** 2020, **17**(4):e1003082.
30. Yepes-Nuñez JJ, Li S-A, Guyatt G, Jack SM, Brozek JL, Beyene J, Murad MH, Rochweg B, Mbuagbaw L, Zhang Y *et al*: **Development of the summary of findings table for network meta-analysis.** *Journal of Clinical Epidemiology* 2019, **115**:1-13.
31. Cipriani A, Higgins JPT, Geddes JR, Salanti G: **Conceptual and technical challenges in network meta-analysis.** *Ann Intern Med* 2013, **159**(2):130-137.
32. Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, Ioannidis JP, Straus S, Thorlund K, Jansen JP *et al*: **The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: checklist and explanations.** *Annals of internal medicine* 2015, **162**(11):777-784.

Figures

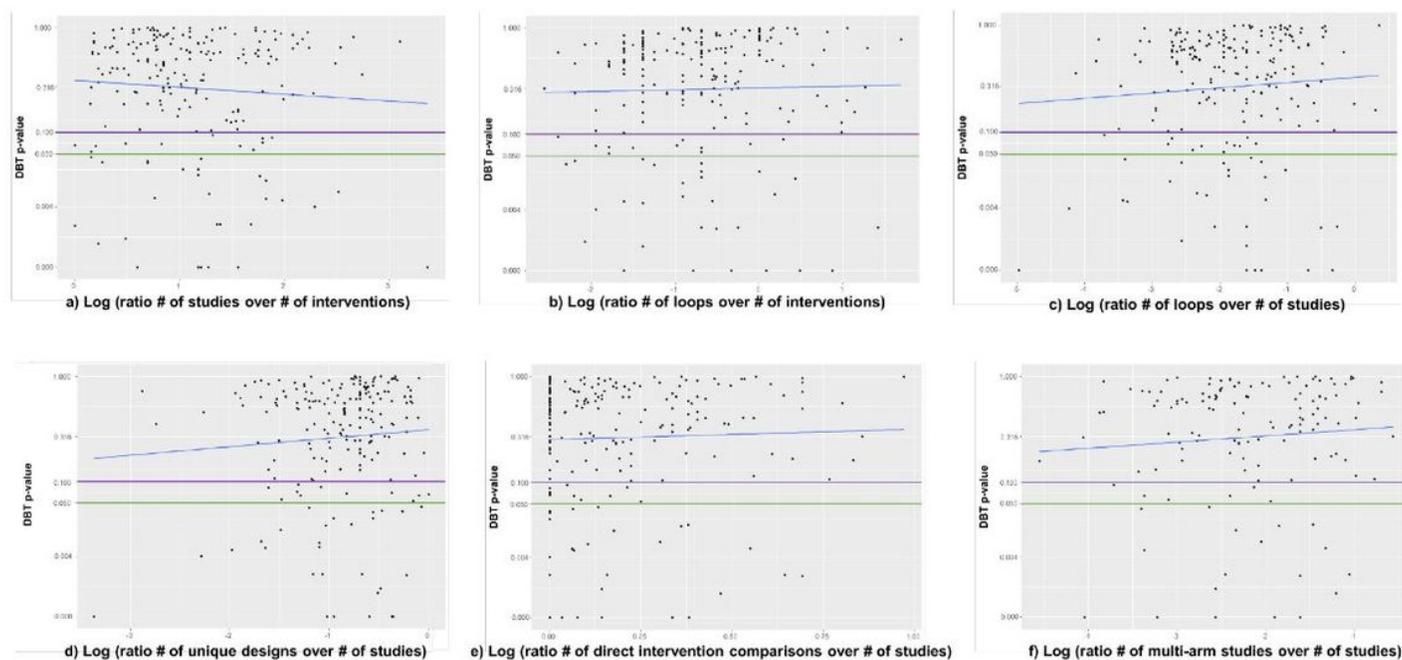


Figure 1

Plot of p-values (fourth-root scale) of the DBT model against ratios of network structural characteristics (logarithmic scale) Panel (a) explores the ratio of the number of studies to the number of interventions in a network (correlation coefficient [p-value] = -0.07 [0.31]); panel (b) explores the ratio of the number of loops to the number of interventions in a network (correlation coefficient [p-value] = 0.02 [0.72]); panel (c)

explores the ratio of the number of loops to the number of studies in a network (correlation coefficient [p-value] = 0.08 [0.28]); panel (d) explores the ratio of the number of unique designs to the number of studies in a network (correlation coefficient [p-value] = 0.07 [0.28]); panel (e) explores the ratio of the number of direct intervention comparisons to the number of studies in a network (correlation coefficient [p-value] = 0.04 [0.60]); panel (f) explores the ratio of the number of multi-arm studies to the number of total studies in a network (correlation coefficient [p-value] = 0.09 [0.19]). All analyses have used the REML estimator for heterogeneity. The horizontal green and purple lines represent the cut-off p-value = 0.05 and p-value = 0.10, respectively. The blue diagonal line is the regression line. * Networks with no multi-arm studies were treated as networks including a single multi-arm study, to avoid excluding them from the plot, since the logarithm could not be calculated. Abbreviations: DBT, design-by-treatment interaction model; REML, restricted maximum likelihood

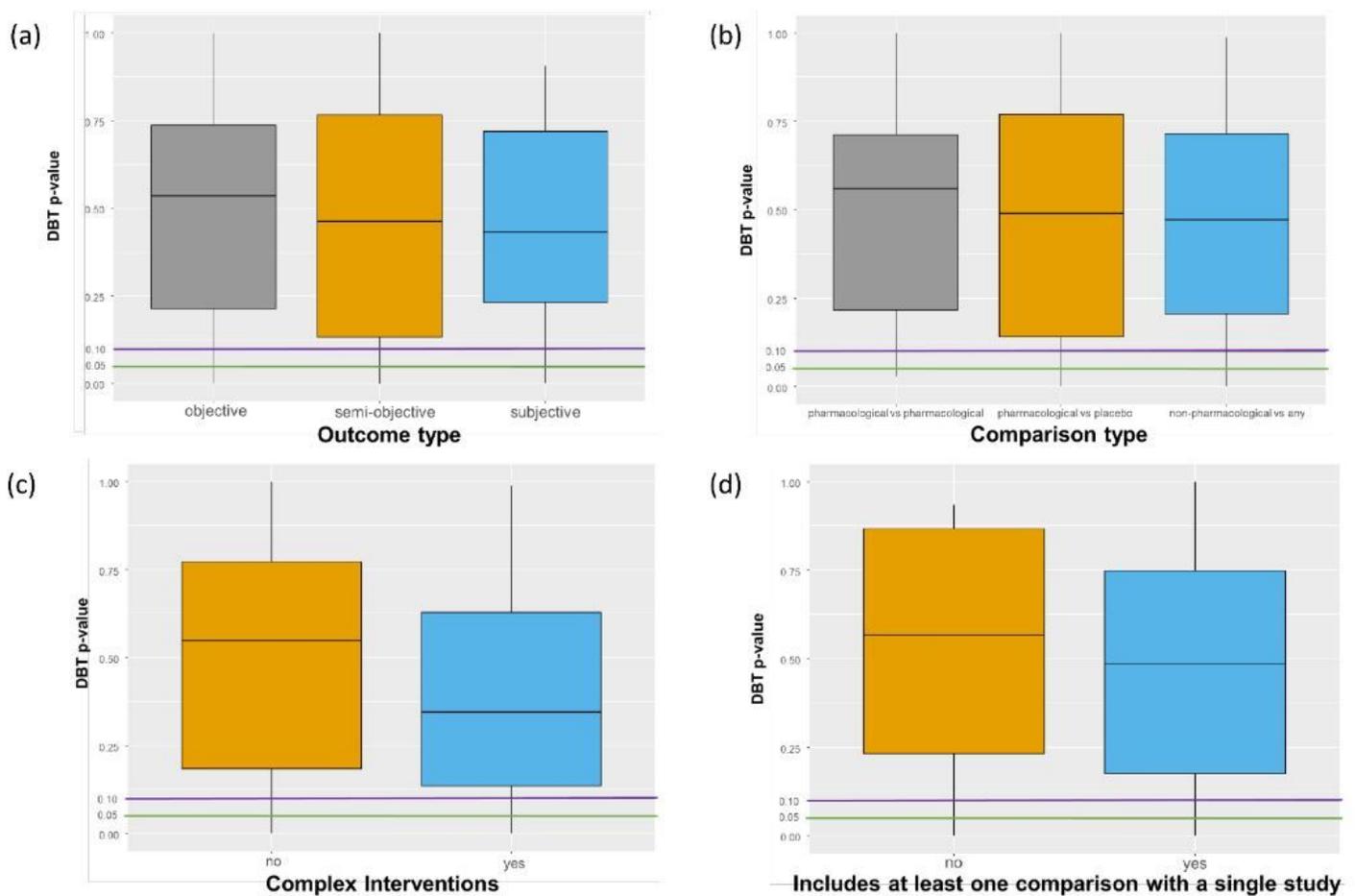


Figure 2

Box plot of p-values (fourth-root scale) of the DBT model per (a) type of outcome, (b) type of intervention comparison, c) presence of at least one direct intervention comparison with a single study, and (d) presence of complex interventions. The horizontal green and purple lines represent the cut-off p-value =

0.05 and p-value = 0.10, respectively. All analyses have used the REML estimator for heterogeneity. Abbreviations: DBT, design-by-treatment interaction model; REML, restricted maximum likelihood

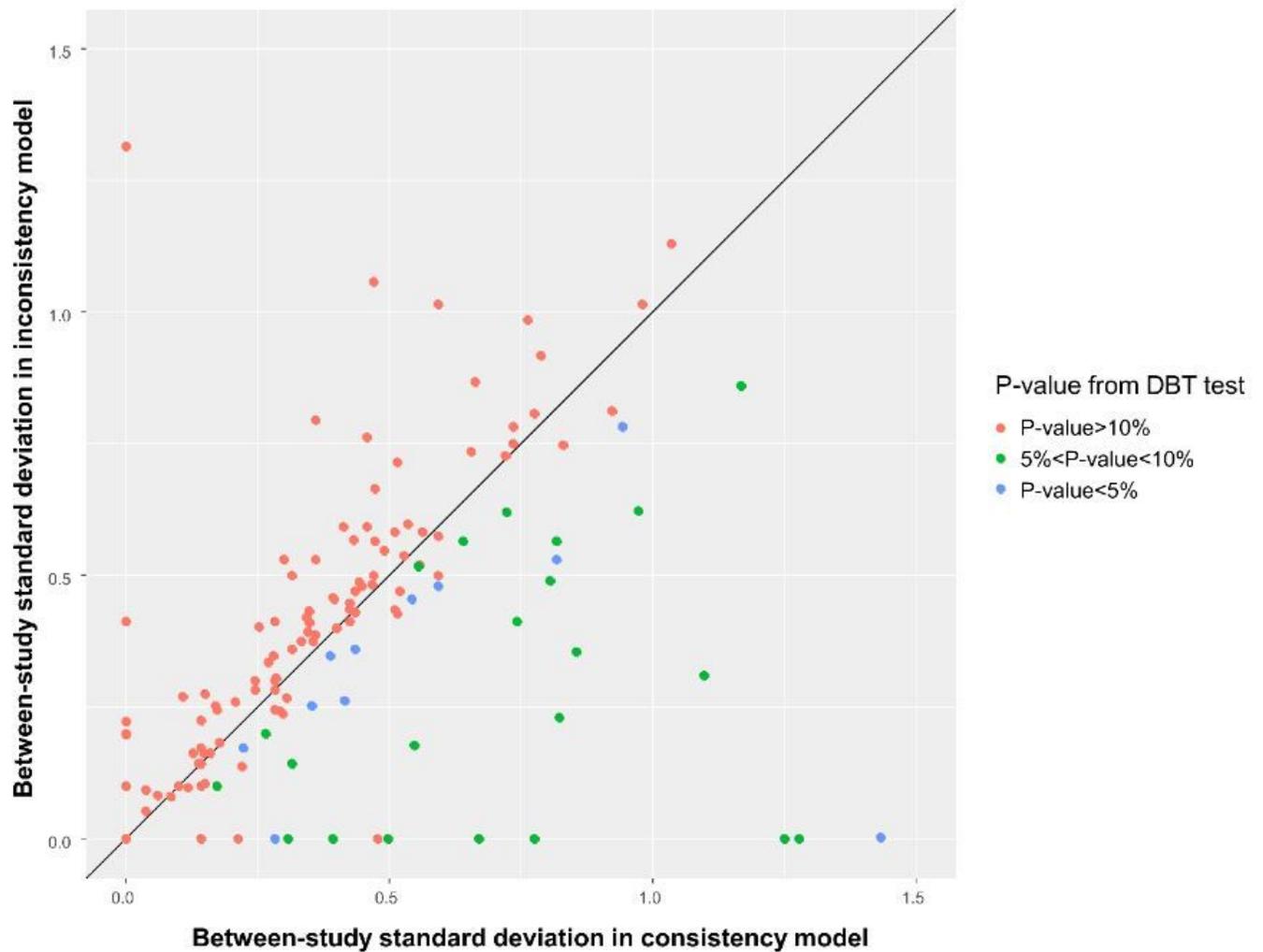


Figure 3

Plot of the between-study standard deviation in consistency against the inconsistency model. The black diagonal line represents equality in between-study standard deviation between consistency and inconsistency models. Red points represent the networks consistent at the threshold of 10%, blue points represent the inconsistent networks at $\alpha=5\%$, and green points represent networks inconsistent between the thresholds 5% and 10%. All analyses have used the REML estimator for heterogeneity. Abbreviations: DBT, design-by-treatment interaction model; REML, restricted maximum likelihood

Supplementary Files

This is a list of supplementary files associated with this preprint. [Click to download.](#)

- [Supplementary07APR2021.pdf](#)