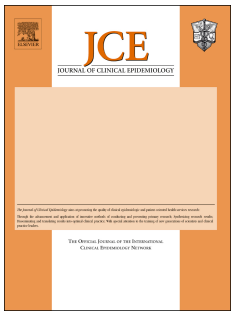
Accepted Manuscript



Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015

Maria Petropoulou, Adriani Nikolakopoulou, Areti-Angeliki Veroniki, Patricia Rios, Afshin Vafaei, Wasifa Zarin, Myrsini Giannatsi, Shannon Sullivan, Andrea C. Tricco, Anna Chaimani, Matthias Egger, Georgia Salanti

PII: S0895-4356(16)30256-6

DOI: 10.1016/j.jclinepi.2016.11.002

Reference: JCE 9271

To appear in: *Journal of Clinical Epidemiology*

Received Date: 25 July 2016

Revised Date: 7 October 2016

Accepted Date: 9 November 2016

Please cite this article as: Petropoulou M, Nikolakopoulou A, Veroniki A-A, Rios P, Vafaei A, Zarin W, Giannatsi M, Sullivan S, Tricco AC, Chaimani A, Egger M, Salanti G, Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015, *Journal of Clinical Epidemiology* (2016), doi: 10.1016/j.jclinepi.2016.11.002.

Bibliographic study showed improving statistical methodology of network meta-analyses published between 1999 and 2015

Maria Petropoulou[1], Adriani Nikolakopoulou[2,1], Areti-Angeliki Veroniki[3], Patricia Rios[3], Afshin Vafaei[3], Wasifa Zarin[3], Myrsini Giannatsi[4], Shannon Sullivan[3], Andrea C. Tricco[3,5], Anna Chaimani[1], Matthias Egger[2], Georgia Salanti[2,1,6]

[1] Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece.

[2] Institute of Social and Preventive Medicine, University of Bern, Switzerland

[3] Li Ka Shing Knowledge Institute, St. Michael's Hospital, 209 Victoria Street, East Building, Toronto, Ontario, M5B 1W8, Canada

[4]Centre for Mental Health and Safety, University of Manchester, UK

[5] Epidemiology Division, Dalla Lana School of Public Health, University of Toronto, 6th floor, 155 College Street, Toronto, Ontario, M5T 3M7, Canada

[6]Bern Institute of Primary Health Care (BIHAM), University of Bern, Switzerland

Correspondence to: Georgia Salanti
Institute of Social and Preventive Medicine (ISPM)
University of Bern
Switzerland

Finkenhubelweg 11
Room 811
3012 Bern
Switzerland
Tel +41 31 631 3555
georgia.salanti@ispm.unibe.ch

Keywords: multiple interventions; meta-epidemiology; mixed-treatment comparisons; inconsistency; reporting

Abstract

Objective: To assess the characteristics and core statistical methodology specific to network meta-analyses (NMAs) in clinical research articles.

Study Design and Setting: We searched Medline, Embase and the Cochrane Database of Systematic Reviews from inception until April 14, 2015 for NMAs of randomized controlled trials (RCTs) including at least four different interventions. Two reviewers independently screened potential studies, while data abstraction was performed by a single reviewer and verified by a second.

Results: A total of 456 NMAs, which included a median (interquartile range) of 21 (13 to 40) studies and 7 (5 to 9) treatment nodes were assessed. A total of 125 NMAs (27%) were star networks; this proportion declined from 100% in 2005 to 19% in 2015 (p=0.01 by test of trend). An increasing number of NMAs discussed transitivity or inconsistency (0% in 2005, 86% in 2015, p<0.01) and 150 (45 %) and used appropriate methods to test for inconsistency (14% in 2006, 74% in 2015, p<0.01). Heterogeneity was explored in 256 NMAs (56%), with no change over time (p=0.10). All pairwise effects were reported in 234 NMAs (51%), with some increase over time (p=0.02). The hierarchy of treatments was presented in 195 NMAs (43%), the probability of being best was most commonly reported (137 NMAs, 70%) but use of SUCRA (surface under the cumulative ranking curves) increased steeply (0% in 2005, 33% in 2015, p<0.01).

Conclusion: Many NMAs published in the medical literature have significant limitations in both the conduct and reporting of the statistical analysis and numerical results. The situation has however improved in recent years, in particular with respect to the evaluation of the underlying assumptions, but considerable room for further improvements remains.

278 words

What is new?

Key findings

Although the amount of evidence (the number of treatments and studies) included in published NMAs remains stable, the undertaking and reporting of statistical methods have significantly improved over the years. The assumptions underlying NMA are increasingly discussed and evaluated using appropriate methods. Less than 10% of NMAs published in 2014 and 2015 failed to evaluate the assumptions of the joint synthesis.

What this adds to what is known

This meta-epidemiological study presents the largest collection of published NMAs over the past 16 years. It provides an overview of the structural characteristics and statistical methodology of 456 published networks of interventions. It shows that the statistical methods in NMA have considerably improved in all aspects and some, such as the use of appropriate methods to evaluate the plausibility of the assumptions, are now routinely performed. We conclude that the increasingly populous community of NMA methodologists is quickly advancing through the learning curve of statistical methods employed in NMA.

What is the implication, what should change now

The updated description of the structural characteristics of the published NMAs can be used to inform pragmatic simulations studies and the development of methods that are relevant to the type of networks typically found in the medical literature.

Future tutorials and training should be focused on improving the methodology and reporting on items that, although they have improved, their prevalence remains low, such as the formal exploration of heterogeneity and inconsistency and the presentation of all pairwise treatment effects.

1 Introduction

Network meta-analysis (NMA) is becoming increasingly popular for evidence synthesis [1–4] with enthusiasts considering NMA as the 'new norm' for comparative effectiveness research [5]. However, empirical studies exploring the characteristics of published NMAs of interventions have raised the need for improving the quality of the application of NMA methods [1,2,6–9]. Concerns about inappropriate applications of NMA methods and inadequate and non-transparent reporting of methods and results have been identified as major issues [10–12]. For instance, Nikolakopoulou et al. found that 68% of the NMAs published by the end of 2012 used either inappropriate or unspecified methods to assess inconsistency, while Bafeta et al. concluded that reporting guidelines are necessary to reduce heterogeneity in presentation of NMA results [1,6].

The importance of empirical evidence in a novel, rapidly evolving methodological field is illustrated by the role such studies played in shaping the methodology for conventional pairwise meta-analysis. The assessment of risk of bias in the included studies [13–15], the magnitude and determinants of heterogeneity [16–18], the relative advantages of different methods to evaluate publication bias and small-study effects, [19–21] and the importance of a comprehensive search for relevant studies [22] are examples of meta-epidemiological studies that have guided the choice of optimal methods. We are aware of only three such meta-epidemiological studies in NMA: Song et al. evaluated the prevalence of inconsistency in networks with three treatments [23,24], Veroniki et al. studied the prevalence of inconsistency in complex NMAs that included at least four treatments using two alternative methods [25] while Chaimani et al. have provided empirical evidence about the impact of risk of bias and small study effects [26].

In recent years, the methodology of NMA has been further refined and many tutorials and guidance papers have been published [12,27–31]. Efthimiou et al., in a recent review of methodological articles published until March 2014, [32] found a steep increase after 2011 in the number of relevant publications. In 2012 and 2013, 83 methodological articles were published compared to 58 articles between 2005-2011. For example, until recently most networks were fitted within a Bayesian framework as hierarchical models [1], but new NMA methods include publications by White et al. suggesting NMAs can be viewed as a specific case of multivariate meta-regression, and by Higgins et al. presenting a new test for inconsistency [33,34]. Their work enables researchers to fit models using frequentist software. Stata routines were made available by White et al. and Chaimani et al. that simplified NMA implementation by non-statisticians [35–39].

In this paper, we aim to describe how methodologies specific to NMA and reporting quality of results has evolved over time, monitor the rate of adoption for the new methodological developments and provide an updated overview of the characteristics of published networks.

2 Methods

2.1 Inclusion criteria

Networks of randomized controlled trials (RCTs) were eligible if they included at least four different interventions (defined as different drugs or other medical treatments, or different schedules, doses or formulations of the same treatment) including placebo, no treatment, waiting list or other control interventions. We excluded networks that included observational or diagnostic test accuracy studies, NMAs in which the number of trials was smaller than the number of interventions and articles that performed naive indirect comparisons by pooling data across study arms.

2.2 Literature search

An expert librarian compiled the literature search, which was peer-reviewed by a second librarian using the Peer Review of Electronic Search Strategies (PRESS) checklist [40]. Subsequently, Medline, Embase and the Cochrane Database of Systematic Reviews were searched from inception until April 14, 2015. The full literature search strategy is provided in the Appendix. There were no language restrictions on our search.

2.3 Screening

After a pilot-test of the eligibility criteria, pairs of reviewers (MG, AC, MP, AAV, PR, AV, SS) independently screened the titles and abstracts from the literature search. Potentially relevant full-text papers were screened in the same manner. Conflicts were resolved by a third reviewer (AAV, AN, AC, GS, PR) to increase consistency.

2.4 Data extraction and data items

One reviewer (MP, AN, AC, AAV, MG, PR, AV, SS or WZ) abstracted data from the included studies and then data was checked by another reviewer. Data items included general publication characteristics: e.g.., year of publication, country of corresponding author and journal of publication. We recorded whether the primary outcome measured efficacy or safety and classified outcomes as dichotomous, continuous, time-to-event or rate. For networks that indicated none or more than one outcome as primary, we used discussion to reach consensus for a final decision. We extracted the total number of compared interventions (termed nodes of the network henceforth), the reference intervention when reported (placebo, usual care, no

treatment, or active treatment). We categorized each network according to the type of treatment comparisons as pharmacological versus placebo, pharmacological versus pharmacological or non-pharmacological versus any treatment (for details see Turner et al. [41] and Nikolakopoulou et al. [1]). When the reference treatment was not explicitly defined, one of the following was selected as the reference treatment node: placebo, usual care, or no treatment. We also categorised the shape of network according to the presence or absence of at least one closed loop.

We only examined methodological characteristics that are specific to NMA methodology, with an emphasis on statistical analysis and reporting. We recorded whether and how the authors evaluated the plausibility of transitivity (the comparability of the distribution of effect modifiers across treatment comparisons) prior to the data synthesis [27]. For networks including at least one closed loop, we also recorded the use of any statistical method to evaluate consistency (e.g. statistical methods such as those described in Donegan et al. (2013)) [42]. We categorized the method used to derive indirect and/or network estimates (e.g. a Bayesian hierarchical model, a multivariate meta-analysis approach), effect measure employed to undertake the analysis (such as odds ratio or mean difference) and whether fixed-effects model, random-effects model or both computational models were used. We also assessed any secondary analyses such as subgroup, network meta-regression, or sensitivity analyses that the authors performed to investigate potential sources of heterogeneity or inconsistency. We examined whether authors assessed small-study effects, whether they considered the potential for publication bias and the methods they applied to evaluate their impact on the results. We recorded whether all possible relative effects were presented (in the main text or as supplementary material) or only a subset of them. We also examined whether the authors presented the estimated hierarchy of the included interventions and which measure was used for this purpose (such as probability of being the best or surface under the cumulative ranking (SUCRA) curve [11,43,44]).

2.5 Statistical analysis

We performed a descriptive analysis for all the characteristics we extracted from the eligible networks of interventions. For characteristics that have been previously identified as needing improvement [1,2,45] – such as the adoption of appropriate methods to evaluate consistency, the uptake of frequentist methods for NMA, and the transparency of reporting – we evaluated whether there was a change in frequency over time. We formally tested the trend using a X^2 test for trends in proportions for dichotomous characteristics and the Cox-Stuart test for trends in continuous characteristics [46]. We also investigated any associations between methods employed and the complexity of the evidence-base such as the number of nodes and the

presence of closed loops. All analyses were performed in R software [47] using the *trend* library [48].

3 Results

The search identified 3727 citations and after screening the titles and abstracts, 877 potentially relevant full text articles were reviewed. In total, 456 NMAs met all inclusion criteria. The full selection process for the included networks is presented in Figure 1.

Figure 1 Flow chart of the selection process for the included networks.

Table 1 presents a description of the sample. The median number of studies per network was 21 (interquartile range (IQR) 13 to 40) and the median number of treatments was 7 with an IQR of 5 to 9. The majority of NMAs included at least one closed loop (331, 73%) while nearly a quarter of the networks (125 NMAs, 27%) did not contain a closed loop of evidence (called star-shaped networks). The majority of the included NMAs had a beneficial primary outcome (260 NMAs, 57%) that was commonly measured using a dichotomous outcome (267 NMAs, 59%). The primary outcome was measured using continuous data in 135 NMAs (30%). Two-thirds of the networks compared pharmacological treatments and included placebo (299 NMAs, 66%) while one in five compared only pharmacological interventions (88 NMAs, 19%). A small number of networks had a mixture of pharmacological, non-pharmacological and control treatments (69 NMAs, 15%).

Table 1 Characteristics of 456 NMAs published until 2015. IQR: Interquartile range.

Most of the included articles were published in general medicine journals (183 NMAs, 40%), the most common was the *British Medical Journal* (28 NMAs, 6%) followed by *Current Medical Research & Opinion* (24 NMAs, 5%). Corresponding authors of 234 NMAs (51%) had an affiliation in a European country (92 NMAs, 20% for United Kingdom) and 140 NMAs (31%) had an affiliation in the United States.

3.1 Time trends in the characteristics of NMAs

Between 1999 and 2004 only 6 NMAs were published (1 in 1999, 2 in 2000, 1 in 2003 and 2 in 2004). The number of NMAs published per year after 2004 is presented in Table 2. It is evident that the number of published studies applying NMA methods to clinical research questions has been increasing significantly over the last two decades (p=0.04).
Overall, the networks of published NMAs do not seem to include more studies (p=0.08, data not shown) or to compare more treatments in recent years (p=0.72, data not shown).

However, the proportion of star-shaped networks has decreased with time while the number of networks with at least one closed loop has increased (p=0.01, Table 2).

Table 2 Networks published between 2005 and 2015 (until 15 April) and their characteristics. The entries in the table show number of networks and respective percentages. P-values are from a trend test. *There are 6 networks published before 2005 and are included in the total NMA group. ** In the test for trend for the total number of published NMAs we excluded the year 2015 as it is not complete. ***Here the denominator is the number of articles with at least one closed loop (number of NMAs published minus the star-shaped NMA).

3.1.1 Evaluation of clinical and statistical assumptions

In three quarters of the included articles (353 NMAs, 77%) we could not identify any statement suggesting that the transitivity (or similarity) assumption was considered or evaluated (Table 2). However, this has improved considerably over time, with 77% of the networks published in 2015 discussing transitivity (p<0.01). In five networks, the authors expressed concerns about the presence of potential non-transitivity and its impact on the results. Among the remaining 100 networks that did report how transitivity was evaluated, the majority compared the study characteristics (76 networks) (Appendix Table 1). One in five of the published networks (22%) reported that the transitivity assumption is likely to hold; the rate of articles reporting this information has increased over time (p<0.01).

Statistical evaluation of consistency was possible for 331 networks that included at least one closed loop. Almost one third of these NMAs (94 NMAs, 28%) did not report any method used for the statistical evaluation of consistency. Nearly half of the networks (150 NMAs, 45%) used appropriate statistical methods to evaluate consistency; their uptake has increased over time (p<0.01, Table 2).

The exact method used for assessing inconsistency in each network is detailed in Appendix Table 2. The loop-specific approach [49] was the most commonly employed method (59 NMAs, 18%), followed by the node-splitting approach [50] (39 NMAs, 12%). The design-by-treatment interaction model [34] that was introduced in 2012 was used in very few networks (5 NMAs, 2%). The proportion of NMAs that considered transitivity or consistency using any method increased significantly over time (p<0.01, Table 2); 86% of the articles published in 2015 reported mentioned at least one of the two terms.

Appendix Table 1 Reporting and evaluation of transitivity. Number of articles and percentages.
Appendix Table 2 Statistical methods used to evaluate the consistency assumption in 331 NMAs with at least one closed loop.

In total we found 76 networks that did not evaluate either transitivity or inconsistency (23% for the networks where assessment of both was possible); the percentage has dropped from 35% in 2010 to 19% in 2012 and to 8% in 2014/2015 ($p<0.01$).

3.1.2 Statistical synthesis of data

We found that the odds ratio (177 NMAs, 39%) and the mean difference (89 NMAs, 20%) were the most frequently used effect sizes for networks with dichotomous and continuous data respectively. It was not always clear whether the random or fixed effects model was used although reporting improved with time ($p=0.01$, Table 2). The majority of the included networks (230 NMAs, 50%) performed the analysis using a random-effects model. Of the 170 networks (37%) that used the fixed-effect model, the majority (141 NMAs, 83%) also applied the random-effects approach either as a sensitivity analysis or with the aim to choose between the two models.

The percentage of articles reporting the statistical method used to fit NMA has increased over time ($p<0.01$, Table 2) and only a small proportion of networks (24 NMAs, 5%) did not report the NMA method used. The Bayesian hierarchical model remains the most popular approach for NMA (302 NMAs, 64%). Only 80 (18%) networks which included at least one multi-arm study employed methods to derive the treatment effect that ignored correlations (e.g. adjusted indirect comparison meta-analysis or Bucher method). NMAs using multivariate meta-analysis or multivariate meta-regression was encountered in only 5 publications while the graph-theoretical method for NMA was employed in one network.

Figure 2 Method used to synthesize the data in relation to the shape of the network. If a network applied more than one method, it is included in all relevant categories.

Figure 2 illustrates that review authors tend to employ a Bayesian hierarchical approach and meta-regression more often in the presence of closed loops.

Approximately half of the networks (256 NMAs, 56%) investigated potential sources of heterogeneity or inconsistency via the use of subgroup, meta-regression or sensitivity analysis. This remained unchanged over time ($p=0.10$). Small-study effects and publication bias were assessed in 143 (31%) networks. Funnel plots (116 NMAs, 81%) and regression tests (82 NMAs, 57%) were the most prevalent methods for the assessment of publication bias and 7 NMAs (5%) applied trim and fill method (Appendix Table 3). These methods were primarily applied to the direct comparisons and only a handful of networks (6 NMAs, 4%) considered the extensions for the aforementioned approaches into the context of NMA, such as the comparison adjusted funnel plot and extended selection models [36,51–53].

Appendix Table 3 Methods employed to assess small-study effects and/or publication bias in 143 NMAs.

3.1.3 Presentation of results

The percentage of NMAs that report outcome data for the primary outcome decreased with time (p=0.03). Half of the included reviews (234 NMAs, 51%) presented all possible relative treatment effects; the other half present only a selective set of comparisons of interest. Newer articles tend to be more inclusive and present all pairwise effects (p=0.02, Table 2). One network (0.2%) did not report any relative treatment effect. The relative hierarchy of the treatments was presented in 43% (195 NMAs) of networks. Probability of being the best was the most popular ranking measure employed to derive a treatment hierarchy (166 NMAs, 85%) followed by SUCRA (39 NMAs, 20%) (Appendix Figure 1). While the frequency of exclusive use of the 'probability of being the best' has not changed significantly over time (p=0.86, Table 2), the use of SUCRAs has increased sharply (p<0.01, Table 2).

Appendix Figure 1 Ranking measures used in the included networks per year (1999-2015).

4 Discussion

We identified 456 NMAs that were published between 1999 and 2015 by searching three bibliographic databases and assessed the characteristics of their statistical analysis and reporting of results. We found that many NMAs published during this time period have significant methodological limitations, but that the application of some methodological elements improved in recent years. For example an increasing number of published NMAs addressed transitivity or inconsistency, by 2015 about three quarters of analyses used appropriate methods to test for inconsistency. The quality and transparency of reporting also increased: in recent years around 90% of articles clearly reported whether a random-effects or fixed-effect model was used, and in 2015 all reports included a description of the statistical methods used. However, important deficiencies in the application of NMA methods remain: discussion of the transitivity assumption was rare and only about half of the articles reported the results of all pairwise comparisons with no increase in recent years. It should be noted that the Bayesian hierarchical model remained the most popular approach for NMA during the study period: only five articles reported the use of frequentist multivariate meta-analysis or meta-regression.

Our study has several strengths. Expert librarians conducted a comprehensive literature search in multiple bibliographic databases and our study team peer reviewed the literature search, screened articles in duplicate using standardized tools, and extracted and verified the relevant information using standardized forms. Calibration exercises among all reviewers preceded

each step of the review. We identified 456 NMAs, which were published over the past 17 years and although our search may have missed some articles, it is very likely that we identified a set of NMAs that accurately reflects current and past practice and methodological development. To the best of our knowledge, this is the largest and most up-to-date collection of published NMAs compiled to date. It includes nearly three times the data included in Bafeta et al. [2], more than twice the data included in Nikolakopoulou et al., [1] and about 40% more data compared to the collection recently published by Chambers et al. [54]. This is also the first study to formally investigate the changes in the methodology and reporting of NMAs over time.

Previous empirical studies have motivated various groups to develop recommendations to improve the practice and reporting of NMAs. For example, Hutton et al. developed an extension of the PRISMA statement (Preferred Reporting Items for Systematic reviews and Meta-analysis) for the reporting of systematic reviews that include NMAs of health care interventions [45]. The National institute of Clinical excellence (NICE) in the UK and the International Society For Pharmacoeconomics and Outcomes Research (ISPOR) have also developed relevant guidance [31,55].

We observed an important improvement in the statistical methodology and reporting of NMAs over time. The PRISMA statement was published only recently and we do not expect it to have had any major impact on the improvements in reporting observed in our study. The educational papers published several years ago [12,27,28,56,57] might, however, have had some impact. Alternatively, improvements in NMA methods might be due to statisticians and clinicians becoming more experienced with NMA techniques. This 'learning curve' could be accelerated by developments making methods more accessible and more widely available. We found that features adding to the complexity of evidence structures, such as closed loops and large numbers of studies, were more common in NMAs performed within a Bayesian framework. The frequentist approaches to NMA that can cope with large and complex networks, such as the routine in Stata that fits the multivariate random-effects meta-analysis model [33,39], have not been applied widely but are expected to be increasingly used in the next years. Similarly, the design-by-treatment interaction model to test for inconsistency [34] that was introduced in 2012 was used in only few networks. This may be explained by the fact that the parameterization of this model is complex and until recently no dedicated software was available [39].

The transitivity assumption is imperative for a valid NMA and also needs to be met when informally comparing effects estimates from separate pairwise meta-analyses. In our large collection of networks, the vast majority of reports did not mention the transitivity assumption and only a handful of NMA reviews assessed it through the comparison of the distribution of

effect modifiers. Suggestions of accompanied detailed protocols on which authors should base their NMA reviews [12,27] have started to be applied. Among other merits, establishment of NMA protocol registration could substantially improve reporting and evaluation of transitivity through the a priori description of approaches that will be used to assess it.

Almost 10 years ago Sutton and Higgins stated that "*time will tell whether this [NMA] is how efficacy of treatments will be routinely compared in the future*" [58]. The substantial increase in published NMAs supports the notion that the rationale, importance, and methodology of comparing multiple treatments simultaneously has now become acknowledged among researchers, national and international health-care institutions [56]. It would be useful if the database used for this study was to be regularly updated to monitor developments in NMA publication and serve as a resource for further empirical research to better define the place of NMA in comparative effectiveness research. Several such projects are ongoing or planned: the quality of the systematic reviews providing data for each of the 456 networks is currently being evaluated, networks are currently being re-analysed to estimate the prevalence of statistical inconsistency, updating previous empirical research [25], and a subset of the included NMAs is being evaluated with respect to their conclusiveness using sequential methods. Additionally, the methodology to compare different doses of the same intervention within NMA is not well developed and an ongoing scoping review plans to describe the approaches currently available in the published literature. Finally, our database could inform simulation studies and scenarios evaluating the performance of various methodological approaches so that simulations reflect the most common encountered circumstances, as was previously done in Veroniki et al. [61].

Acknowledgements: We would like to thank Inthuja Selvaratnam and Katerina Vasiliadou for helping with contacting authors and data extraction, respectively.
Georgia Salanti has been supported by a Horizon 2020 Marie-Curie Individual Fellowship (Grant No 703254). Areti Angeliki Veroniki is funded by the Canadian Institutes of Health Research (CIHR) Banting Postdoctoral Fellowship Program. Andrea C. Tricco is supported by a Drug Safety and Effectiveness Network CIHR New Investigator Award in Knowledge Synthesis.

5 References

[1] Nikolakopoulou A, Chaimani A, Veroniki AA, Vasiliadis HS, Schmid CH, Salanti G. Characteristics of Networks of Interventions: A Description of a Database of 186 Published Networks. PLOS ONE 2014;9:e86754. doi:10.1371/journal.pone.0086754.

[2] 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. doi:10.1136/bmj.f3675.

[3] Lee AW. Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. J Clin Epidemiol 2014;67:138–43. doi:10.1016/j.jclinepi.2013.07.014.

[4] Achana F, Hubbard S, Sutton A, Kendrick D, Cooper N. An exploration of synthesis methods in public health evaluations of interventions concludes that the use of modern statistical methods would be beneficial. J Clin Epidemiol 2014. doi:10.1016/j.jclinepi.2013.09.018.

[5] Higgins JPT, Welton NJ. Network meta-analysis: a norm for comparative effectiveness? Lancet Lond Engl 2015;386:628–30. doi:10.1016/S0140-6736(15)61478-7.

[6] Bafeta A, Trinquart L, Seror R, Ravaud P. Reporting of results from network meta-analyses: methodological systematic review. BMJ 2014;348:g1741. doi:10.1136/bmj.g1741.

[7] Song F, Loke YK, Walsh T, Glenny A-M, Eastwood AJ, Altman DG. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ 2009;338:b1147. doi:10.1136/bmj.b1147.

[8] Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R, et al. Indirect comparisons of competing interventions. Health Technol Assess Winch Engl 2005;9:1–134, iii–iv.

[9] Donegan S, Williamson P, Gamble C, Tudur-Smith C. Indirect Comparisons: A Review of Reporting and Methodological Quality. PLOS ONE 2010;5:e11054. doi:10.1371/journal.pone.0011054.

[10] Caldwell DM, Gibb DM, Ades AE. Validity of indirect comparisons in meta-analysis. Lancet 2007;369:270. doi:10.1016/S0140-6736(07)60138-X.

[11] Mavridis D, Giannatsi M, Cipriani A, Salanti G. A primer on network meta-analysis with emphasis on mental health. Evid Based Ment Health 2015;18:40–6. doi:10.1136/eb-2015-102088.

[12] Cipriani A, Higgins JPT, Geddes JR, Salanti G. Conceptual and Technical Challenges in Network Meta-analysis. Ann Intern Med 2013;159:130–7. doi:10.7326/0003-4819-159-2-201307160-00008.

[13] Wood L, Egger M, Gluud LL, Schulz KF, Jüni P, Altman DG, et al. Empirical evidence of bias in treatment effect estimates in controlled trials with different interventions and outcomes: meta-epidemiological study. BMJ 2008;336:601–5. doi:10.1136/bmj.39465.451748.AD.

[14] Savović J, Jones HE, Altman DG, Harris RJ, Jüni P, Pildal J, et al. Influence of reported study design characteristics on intervention effect estimates from randomized, controlled trials. Ann Intern Med 2012;157:429–38. doi:10.7326/0003-4819-157-6-201209180-00537.

[15] Sterne JAC, Jüni P, Schulz KF, Altman DG, Bartlett C, Egger M. Statistical methods for assessing the influence of study characteristics on treatment effects in "meta-epidemiological" research. Stat Med 2002;21:1513–24. doi:10.1002/sim.1184.

[16] Rhodes KM, Turner RM, Higgins JPT. Empirical evidence about inconsistency among studies in a pair-wise meta-analysis. Res Synth Methods 2015. doi:10.1002/jrsm.1193.

[17] Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. Int J Epidemiol 2012;41:818–27. doi:10.1093/ije/dys041.

[18] Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327:557–60. doi:10.1136/bmj.327.7414.557.

[19] Nüesch E, Trelle S, Reichenbach S, Rutjes AWS, Tschannen B, Altman DG, et al. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. BMJ 2010;341:c3515.
[20] Kjaergard LL, Villumsen J, Gluud C. Reported methodologic quality and discrepancies between large and small randomized trials in meta-analyses. Ann Intern Med 2001;135:982–9.
[21] Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. BMJ 2013;346:f2304.
[22] Egger M, Juni P, Bartlett C, Holenstein F, Sterne J. How important are comprehensive literature searches and the assessment of trial quality in systematic reviews? Empirical study. Health Technol Assess Winch Engl 2003;7:1–76.
[23] Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. BMJ 2003;326:472. doi:10.1136/bmj.326.7387.472.
[24] Song F, Xiong T, Parekh-Bhurke S, Loke YK, Sutton AJ, Eastwood AJ, et al. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study. BMJ 2011;343:d4909.
[25] Veroniki AA, Vasiliadis HS, Higgins JP, Salanti G. Evaluation of inconsistency in networks of interventions. Int J Epidemiol 2013;42:332–45. doi:10.1093/ije/dys222.
[26] Chaimani A, Vasiliadis HS, Pandis N, Schmid CH, Welton NJ, Salanti G. Effects of study precision and risk of bias in networks of interventions: a network meta-epidemiological study. Int J Epidemiol 2013;42:1120–31. doi:10.1093/ije/dyt074.
[27] 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;3:80–97. doi:10.1002/jrsm.1037.
[28] 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. doi:10.1186/1741-7015-11-159.
[29] Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. Med Decis Making 2013;33:607–17. doi:10.1177/0272989X12458724.
[30] Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health 2011;14:417–28. doi:10.1016/j.jval.2011.04.002.
[31] Jansen JP, Trikalinos T, Cappelleri JC, Daw J, Andes S, Eldessouki R, et al. Indirect treatment comparison/network meta-analysis study questionnaire to assess relevance and credibility to inform health care decision making: an ISPOR-AMCP-NPC Good Practice Task Force report. Value Health 2014;17:157–73. doi:10.1016/j.jval.2014.01.004.
[32] Efthimiou O, Debray TPA, van Valkenhoef G, Trelle S, Panayidou K, Moons KGM, et al. GetReal in network meta-analysis: a review of the methodology. Res Synth Methods 2016. doi:10.1002/jrsm.1195.
[33] 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;3:111–25.
[34] Higgins JPT, Jackson D, Barrett JK, Lu G, Ades AE, White IR. Consistency and insconsistency in network meta-analysis: concepts and models for multi-arm studies. Res Synth Methods 2012;3:98–110.
[35] van Valkenhoef G, Lu G, de Brock B, Hillege H, Ades AE, Welton NJ. Automating network meta-analysis. Res Synth Methods 2012;3:285–99. doi:10.1002/jrsm.1054.
[36] Chaimani A, Salanti G. Visualizing assumptions and results in network meta-analysis: The network graphs package. Stata J 2015;15:905–50.

[37] Chaimani A, Higgins JPT, Mavridis D, Spyridonos P, Salanti G. Graphical tools for network meta-analysis in STATA. PloS One 2013;8:e76654. doi:10.1371/journal.pone.0076654.
[38] Gerta Rücker, Guido Schwarzer, Ulrike Krahn and Jochem König (2015). netmeta: Network Meta-Analysis using Frequentist Methods. R package version 0.8-0. https://CRAN.R-project.org/package=netmeta
[39] White IR. Network meta-analysis. Stata J 2015;15:951–85.
[40] McGowan J, Sampson M, Salzwedel DM, Cogo E, Foerster V, Lefebvre C. PRESS Peer Review of Electronic Search Strategies: 2015 Guideline Statement. J Clin Epidemiol 2016. doi:10.1016/j.jclinepi.2016.01.021.
[41] Turner RM, Davey J, Clarke MJ, Thompson SG, Higgins JP. Predicting the extent of heterogeneity in meta-analysis, using empirical data from the Cochrane Database of Systematic Reviews. Int J Epidemiol 2012;41:818–27. doi:10.1093/ije/dys041.
[42] Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing key assumptions of network meta-analysis: a review of methods. Res Synth Methods 2013:n/a–n/a. doi:10.1002/jrsm.1085.
[43] Rücker G, Schwarzer G. Ranking treatments in frequentist network meta-analysis works without resampling methods. BMC Med Res Methodol 2015;15:58. doi:10.1186/s12874-015-0060-8.
[44] Salanti G, Ades AE, Ioannidis JP. Graphical methods and numerical summaries for presenting results from multiple-treatment meta-analysis: an overview and tutorial. J Clin Epidemiol 2011;64:163–71. doi:10.1016/j.jclinepi.2010.03.016.
[45] Hutton B, Salanti G, Caldwell DM, Chaimani A, Schmid CH, Cameron C, 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. doi:10.7326/M14-2385.
[46] Cox DR, Stuart A. Some Quick Sign Tests for Trend in Location and Dispersion. Biometrika 1955;42:80–95. doi:10.1093/biomet/42.1-2.80.
[47] R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org. 2014.
[48] Pohlert T. trend: Non-Parametric Trend Tests and Change-Point Detection. R package version 0.2.0. https://CRAN.R-project.org/package=trend
[49] 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;50:683–91.
[50] Dias S, Welton NJ, Caldwell DM, Ades AE. Checking consistency in mixed treatment comparison meta-analysis. Stat Med 2010;29:932–44. doi:10.1002/sim.3767.
[51] Chaimani A, Salanti G. Using network meta-analysis to evaluate the existence of small-study effects in a network of interventions. Res Synth Methods 2012;3:161–76.
[52] Mavridis D, Efthimiou O, Leucht S, Salanti G. Publication bias and small-study effects magnified effectiveness of antipsychotics but their relative ranking remained invariant. J Clin Epidemiol 2015. doi:10.1016/j.jclinepi.2015.05.027.
[53] Mavridis D, Sutton A, Cipriani A, Salanti G. A fully Bayesian application of the Copas selection model for publication bias extended to network meta-analysis. Stat Med 2013;32:51–66. doi:10.1002/sim.5494.
[54] Chambers JD, Naci H, Wouters OJ, Pyo J, Gunjal S, Kennedy IR, et al. An assessment of the methodological quality of published network meta-analyses: a systematic review. PLOS ONE 2015;10:e0121715. doi:10.1371/journal.pone.0121715.
[55] Ades AE, Caldwell DM, Reken S, Welton NJ, Sutton AJ, Dias S. Evidence synthesis for decision making 7: a reviewer's checklist. Med Decis Making 2013;33:679–91. doi:10.1177/0272989X13485156.
[56] Li T, Puhan MA, Vedula SS, Singh S et al. Network meta-analysis-highly attractive but more methodological research is needed. BMC Med 2011;9:79. doi:10.1186/1741-7015-9-79.

[57] 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 Mak Int J Soc Med Decis Mak 2013;33:641–56. doi:10.1177/0272989X12455847.

[58] Sutton AJ, Higgins JPT. Recent developments in meta-analysis. Stat Med 2008;27:625–50. doi:10.1002/sim.2934.

[59] The guidelines manual | 1 Introduction | Guidance and guidelines | NICE n.d. http://www.nice.org.uk/article/pmg6/chapter/1%20introduction (accessed April 29, 2015).

[60] Jansen JP, Fleurence R, Devine B, Itzler R, Barrett A, Hawkins N, et al. Interpreting indirect treatment comparisons and network meta-analysis for health-care decision making: report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: part 1. Value Health 2011;14:417–28. doi:10.1016/j.jval.2011.04.002.

[61] Veroniki AA, Mavridis D, Higgins JP, 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. doi:10.1186/1471-2288-14-106.

Table 1 Characteristics of 456 NMAs published until 2015. IQR: Interquartile range.

	Median	***IQR***
Median number of treatments compared	7	(5, 9)
Median number of studies included	21	(13, 40)
	Number of NMAs	***%***
Number of star-shaped networks	125	(27%)
Outcome characteristics		
Beneficial outcome	260	(57%)
Measured as dichotomous	267	(59%)
Measured as continuous	135	(30%)
Compare pharmacological treatments and placebo	299	(66%)
Compare only pharmacological treatments	88	(19%)
Published in general medicine journals[1]	183	(40%)
Published in health services research journals[2]	56	(12%)
Published in specialty journals	217	(48%)
Corresponding author with affiliation in Europe	234	(51%)
Corresponding author with affiliation in USA	140	(31%)

1 Includes the categories Medicine, General & Internal, Pharmacology & Pharmacy, Multidisciplinary Sciences, Medicine, Research & Experimental, Primary Health Care.

2 Includes the categories Health Care Sciences & Services, Health Policy & Services

*Table 2 Networks published between 2005 and 2015 (until 15 April) and their characteristics. The entries in the table show number of networks and respective percentages. P-values are from a trend test. *There are 6 networks published before 2005 and are included in the total NMA group. ** In the test for trend for the total number of published NMA we excluded the year 2015 has it is not complete. ***Here the denominator is the number of articles with at least one closed loop (number of NMAs published minus the star-shaped NMAs).*

Characteristics of NMAs	Total	2005		2006		2007		2008		2009		2010		2011		2012		2013		2014		2015		p-value
Characteristics of the evidence-base																								
Star Networks	125	6	100%	5	42%	2	22%	6	50%	7	26%	7	23%	14	26%	17	29%	19	20%	32	31%	8	19%	**0.01**
Compare pharmacological vs pharmacological	88	1	17%	2	17%	2	22%	3	25%	8	30%	3	10%	5	9%	5	8%	23	24%	29	28%	7	16%	0.15
Compare pharmacological vs placebo	299	5	83%	8	67%	6	67%	9	75%	14	52%	22	73%	43	81%	42	71%	62	65%	56	54%	26	60%	0.31
Compare non-plarmacological vs any	69	0	0%	2	17%	1	11%	0	0%	5	19%	5	17%	5	9%	12	20%	11	11%	18	17%	10	23%	0.05
Evaluation of clinical and statistical assumptions																								
No information or discussion on transitivity	353	6	100%	12	100%	7	78%	11	92%	23	85%	26	87%	46	87%	46	78%	67	70%	71	69%	33	77%	**<0.01**
Reported that transitivity is likely to hold	98	0	0%	0	0%	1	11%	1	8%	4	15%	4	13%	7	13%	13	22%	27	28%	30	29%	10	23%	**<0.01**
Use appropriate methods to test for inconsistency* **	150	NA	NA	1	14%	2	29%	2	33%	6	30%	4	17%	13	33%	16	38%	43	56%	36	51%	26	74%	**<0.01**
Discuss transitivity or inconsistency (at least one of the two)	*285*	0	0%	2	17%	3	33%	5	42%	12	44%	17	57%	30	57%	40	68%	66	69%	72	70%	37	86%	***<0.01***
Statistical synthesis of data																								
Clearly reported whether random or fixed effects are used	400	5	83%	10	83%	7	78%	10	83%	20	74%	25	83%	44	83%	53	90%	91	95%	93	90%	38	88%	**0.01**
Method for NMA reported	432	4	67%	8	67%	9	100%	11	92%	23	85%	30	100%	51	96%	56	95%	95	99%	99	96%	43	100%	**<0.01**
Use Bayesian hierarchical model to fit NMA	302	1	17%	3	25%	3	33%	4	33%	13	48%	19	63%	35	66%	43	73%	77	80%	71	69%	33	77%	**<0.01**
Formal exploration of heterogeneity	256	2	33%	9	75%	5	56%	6	50%	16	59%	20	67%	36	68%	32	54%	56	58%	51	50%	20	47%	0.1
Presentation of results																								
All pairwise effects are presented	234	1	17%	3	25%	2	22%	4	33%	15	56%	17	57%	31	58%	29	49%	54	56%	55	53%	23	53%	**0.02**
Available outcome data	308	4	67%	8	67%	8	89%	10	83%	23	85%	24	80%	36	68%	38	64%	55	57%	71	69%	27	63%	**0.03**
Use Pbest as the only method of ranking	137	1	17%	2	17%	3	33%	1	8%	10	37%	13	43%	16	30%	20	34%	33	34%	32	31%	6	14%	0.86
Use SUCRA to rank treatments	39	0	0%	0	0%	0	0%	0	0%	0	0%	0	0%	1	2%	4	7%	10	10%	9	9%	14	33%	**<0.01**
Number of NMAs published	***456****	***6***		***12***		***9***		***12***		***27***		***30***		***53***		***59***		***96***		***103***		***43***		*0.04***

Appendix Table 1. Reporting and evaluation of transitivity: number of articles and percentages.		
Reported concerns about potential intransitivity		**5 (1.1%)**
Nothing relevant reported		**353 (77.4 %)**
Transitivity established		**100* (21.9 %)**
Method to evaluate transitivity among those claiming transitivity*	Compare the distribution of the effect modifiers across studies grouped by comparison	11 (11%)
	Compare the characteristics of the included trials	76 (76%)
	Use of meta-regression	9 (9%)
	Assumed on the outset**	4 (4%)

*Two networks reported more than one approach to evaluate transitivity**,** the denominator is the total number of networks (456). **Authors stated that they assume on the outset that the transitivity assumption is likely to hold without justifying this decision.

Appendix Table 2. Statistical methods used to evaluate the consistency assumption in 331 NMAs with at least one closed loop.

Inconsistency methods		**Number of networks***
Appropriate methods		164 (49.5 %)
*Appropriate method used**	Loop-specific [1]	59 (17.8 %)
	Node-splitting [2]	39 (11.8 %)
	Back-calculation [2]	8 (2.4 %)
	Comparisons of model fit and parsimony (DIC) [3]	21 (6.3%)
	Inconsistency models (Lumley model [4], Lu and Ades model [5], Design-by treatment model [6])	33 (10 %)
	I^2 [7]	3 (0.9 %)
	Generalized Cochran Q [8]	1 (0.3%)
Inappropriate methods		88 (26.5 %)
Inappropriate methods	Comparison of NMA results with direct meta-analysis results	65 (19.6 %)
	Comparison of the results with previously published results	23 (6.9 %)
	No method reported	94 (28.3 %)

*13 networks used more than one method to evaluate consistency. Here the denominator is the number of networks with at least one closed loop (331).

Appendix Table 3 Methods employed to assess small-study effects and/or publication bias in 143 NMAs.

Method	**Number of networks***
Funnel plots (standard or contour-enhanced [9], [10])	116 (81.1 %)
Regression tests (Egger's [9], Begg's [11], Harbord's [12], Peter's [13])	82 (57.3 %)
Trim and Fill method [14]	7 (4.8 %)
Comparison-adjusted funnel plots [15]	5 (3.4 %)
Extended selection models in NMA [16]–[18]	1 (0.6 %)
L' Abbe plot [19]	1 (0.6 %)
Fail-safe number (Nfs) (Number of studies with zero effect size that would be needed to increase the p-value for the meta-analysis to above 0.05 [20])	2 (1.3 %)
Tolerance level for future null results (Number of studies averaging null results that must be in the file drawers before the overall probability of a Type I error is brought to any desired level of significance [20])	1 (0.6 %)
Other (e.g. analyses the publication rates of the included studies; investigated from previous completed reviews; comparison the results with other study)	4 (2.7 %)

*55 networks applied more than one method. Here the denominator is the number of networks which assess small-study effects and/or publication bias (143).

References

[1] H. C. Bucher, G. H. Guyatt, L. E. Griffith, and S. D. Walter, "The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials," *J. Clin. Epidemiol.*, vol. 50, no. 6, pp. 683–691, Jun. 1997.

[2] S. Dias, N. J. Welton, D. M. Caldwell, and A. E. Ades, "Checking consistency in mixed treatment comparison meta-analysis," *Stat. Med.*, vol. 29, no. 7–8, pp. 932–944, Mar. 2010.

[3] S. Dias, N. J. Welton, A. J. Sutton, D. M. Caldwell, G. Lu, and A. E. Ades, "NICE DSU Technical Support Document 4: Inconsistency in Networks of Evidence Based on Randomised Controlled Trials.," 2011.

[4] T. Lumley, "Network meta-analysis for indirect treatment comparisons," *Stat. Med.*, vol. 21, no. 16, pp. 2313–2324, Aug. 2002.

[5] A. E. A. Guobing Lu, "Assessing evidence inconsistency in mixed treatment comparisons. J Am Statist Assoc 101: 447-459," *J. Am. Stat. Assoc.*, vol. 101, no. June, pp. 447–459, 2006.

[6] J. P. T. Higgins, D. Jackson, J. K. Barrett, G. Lu, A. E. Ades, and I. R. White, "Consistency and inconsistency in network meta-analysis: concepts and models for multi-arm studies," *Res. Synth. Methods*, vol. 3, no. 2, pp. 98–110, Jun. 2012.

[7] D. Jackson, J. K. Barrett, S. Rice, I. R. White, and J. P. T. Higgins, "A design-by-treatment interaction model for network meta-analysis with random inconsistency effects," *Stat. Med.*, vol. 33, no. 21, pp. 3639–3654, Sep. 2014.

[8] U. Krahn, H. Binder, and J. König, "A graphical tool for locating inconsistency in network meta-analyses," *BMC Med. Res. Methodol.*, vol. 13, p. 35, 2013.

[9] M. Egger, G. Davey Smith, M. Schneider, and C. Minder, "Bias in meta-analysis detected by a simple, graphical test," *BMJ*, vol. 315, no. 7109, pp. 629–634, Sep. 1997.

[10] J. L. Peters, A. J. Sutton, D. R. Jones, K. R. Abrams, and L. Rushton, "Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry," *J. Clin. Epidemiol.*, vol. 61, no. 10, pp. 991–996, Oct. 2008.

[11] C. B. Begg and M. Mazumdar, "Operating characteristics of a rank correlation test for publication bias," *Biometrics*, vol. 50, no. 4, pp. 1088–1101, Dec. 1994.

[12] R. M. Harbord, R. J. Harris, and J. A. C. Sterne, "Updated tests for small-study effects in meta-analyses," *Stata J.*, vol. 9 (2), 2009.

[13] J. L. Peters, A. J. Sutton, D. R. Jones, K. R. Abrams, and L. Rushton, "Comparison of two methods to detect publication bias in meta-analysis," *JAMA*, vol. 295, no. 6, pp. 676–680, Feb. 2006.

[14] S. Duval and R. Tweedie, "Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis," *Biometrics*, vol. 56, no. 2, pp. 455–463, Jun. 2000.

[15] A. Chaimani, J. P. T. Higgins, D. Mavridis, P. Spyridonos, and G. Salanti, "Graphical Tools for Network Meta-Analysis in STATA," *PLoS ONE*, vol. 8, p. e76654, Oct. 2013.

[16] L. Trinquart, G. Chatellier, and P. Ravaud, "Adjustment for reporting bias in network meta-analysis of antidepressant trials," *BMC Med. Res. Methodol.*, vol. 12, p. 150, 2012.

[17] A. J. Sutton, F. Song, S. M. Gilbody, and K. R. Abrams, "Modelling publication bias in meta-analysis: a review," *Stat. Methods Med. Res.*, vol. 9, no. 5, pp. 421–445, Oct. 2000.

[18] F. Song, A. J. Eastwood, S. Gilbody, L. Duley, and A. J. Sutton, “Publication and related biases,” *Health Technol. Assess. Winch. Engl.*, vol. 4, no. 10, pp. 1–115, 2000.
[19] K. A. L’ABBÉ, A. S. DETSKY, and K. O’ROURKE, “Meta-Analysis in Clinical Research,” *Ann. Intern. Med.*, vol. 107, no. 2, pp. 224–233, Aug. 1987.
[20] R. Rosenthal, “The file drawer problem and tolerance for null results,” *Psychol. Bull.*, vol. 86, no. 3, pp. 638–641, 1979.

Figure 1

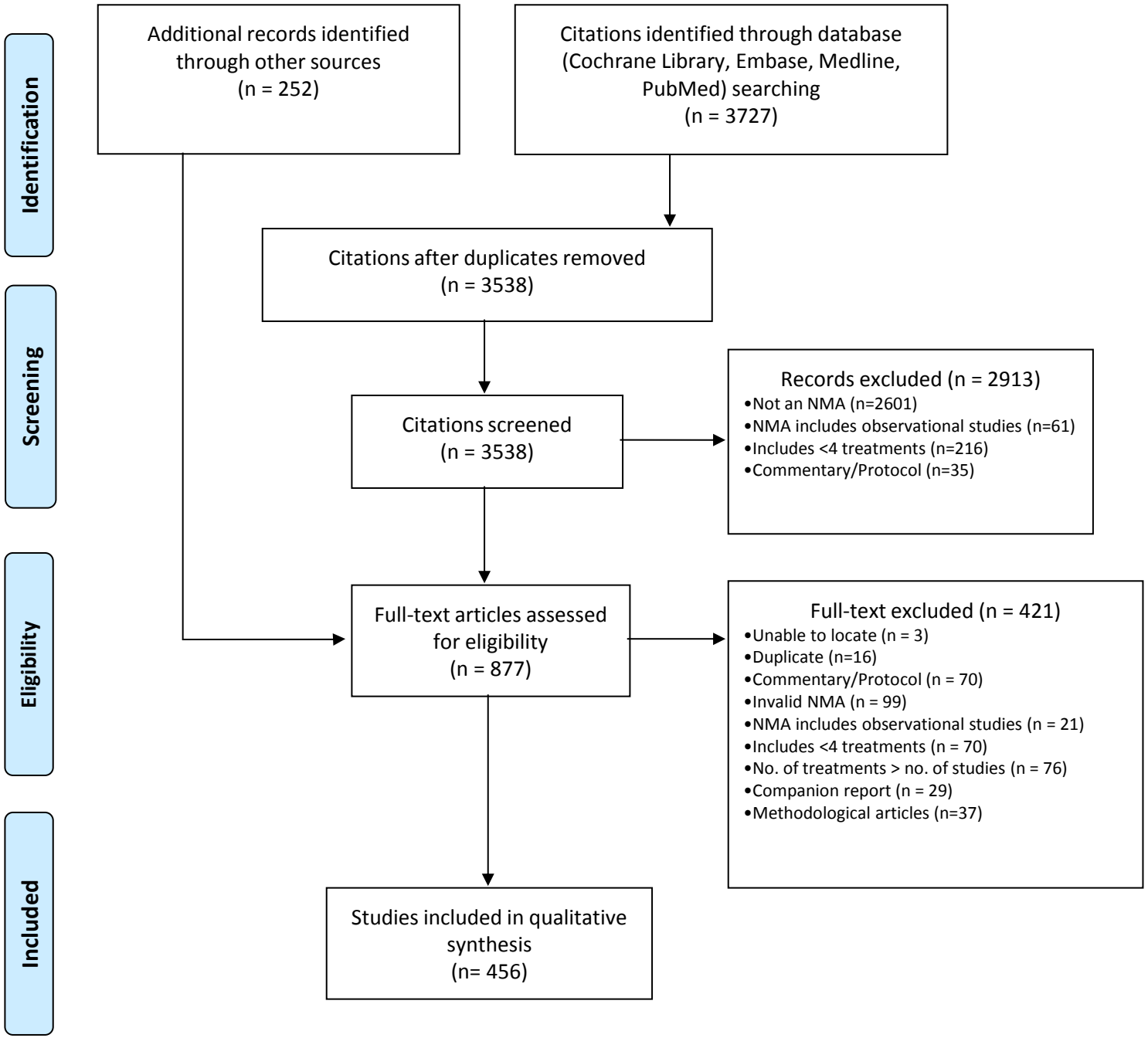
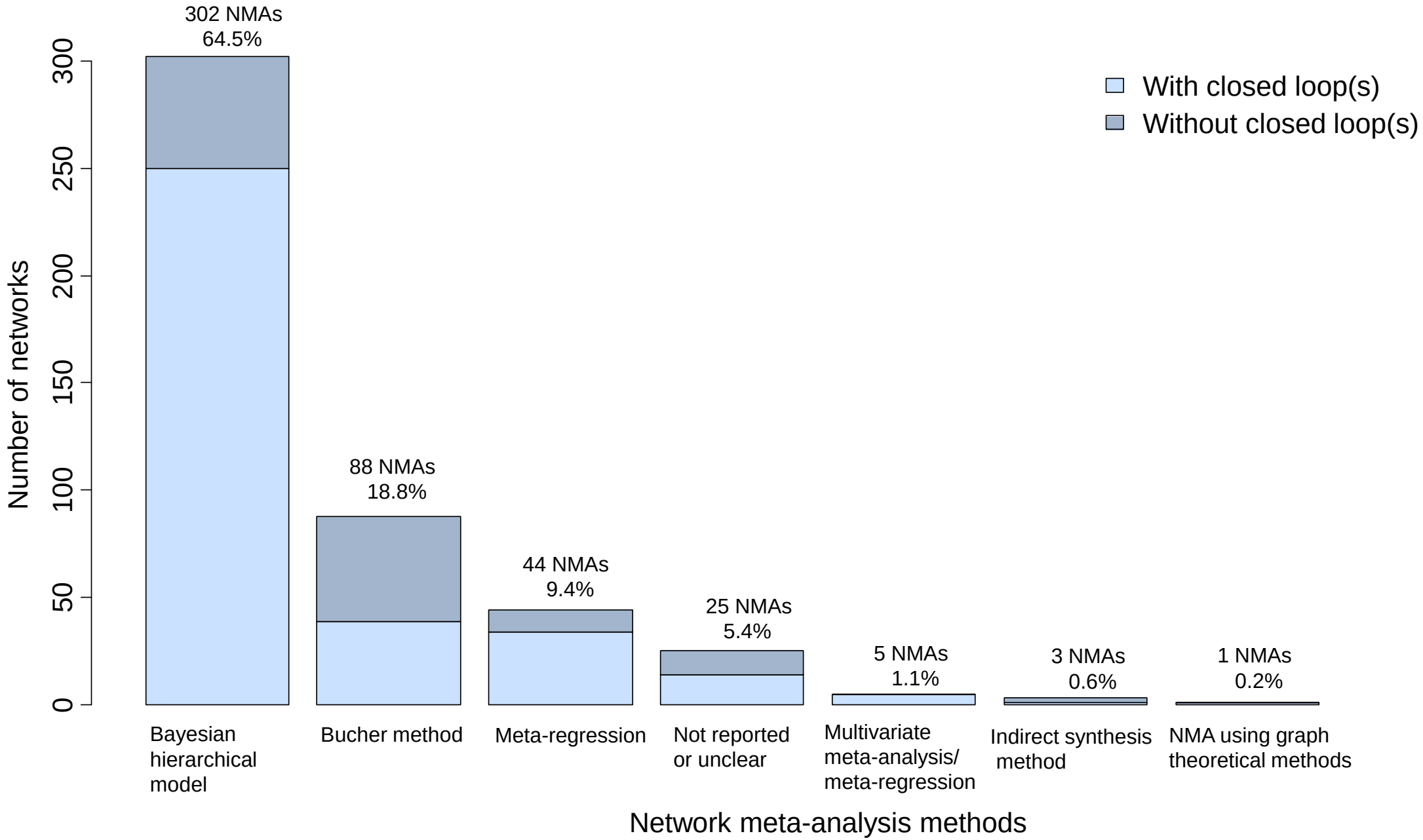

Identification
Screening
Eligibility
Included
Additional records identified through other sources (n = 252)
Citations identified through database (Cochrane Library, Embase, Medline, PubMed) searching (n = 3727)
Citations after duplicates removed (n = 3538)
Citations screened (n = 3538)
Records excluded (n = 2913)
•Not an NMA (n=2601)
•NMA includes observational studies (n=61)
•Includes <4 treatments (n=216)
•Commentary/Protocol (n=35)
Full-text articles assessed for eligibility (n = 877)
Full-text excluded (n = 421)
•Unable to locate (n = 3)
•Duplicate (n=16)
•Commentary/Protocol (n = 70)
•Invalid NMA (n = 99)
•NMA includes observational studies (n = 21)
•Includes <4 treatments (n = 70)
•No. of treatments > no. of studies (n = 76)
•Companion report (n = 29)
•Methodological articles (n=37)
Studies included in qualitative synthesis (n= 456)

Figure 2


302 NMAs
64.5%
88 NMAs
18.8%
44 NMAs
9.4%
25 NMAs
5.4%
5 NMAs
1.1%
3 NMAs
0.6%
1 NMAs
0.2%
With closed loop(s)
Without closed loop(s)
Number of networks
0
50
100
150
200
250
300
Bayesian hierarchical model
Bucher method
Meta-regression
Not reported or unclear
Multivariate meta-analysis/ meta-regression
Indirect synthesis method
NMA using graph theoretical methods
Network meta-analysis methods

Appendix Figure 1