

Integration of evidence from multiple meta-analyses: a primer on umbrella reviews, treatment networks and multiple treatments meta-analyses

John P.A. Ioannidis MD

Meta-analysis is an important research design for appraising evidence and guiding medical practice and health policy.¹ Meta-analyses draw strength from combining data from many studies. However, even if perfectly done with perfect data, a single meta-analysis that addresses 1 treatment comparison for 1 outcome may offer a short-sighted view of the evidence. This may suffice for decision-making if there is only 1 treatment choice for this condition and only 1 outcome of interest and research results are perfect. However, usually there are many treatments to choose from, many outcomes to consider and research is imperfect. For example, there are 68 antidepressant drugs to choose from,² dozens of scales to measure depression outcomes, and biases abound in research about antidepressants.^{3,4}

Given this complexity, one has to consider what alternative treatments are available and what their effects are on various beneficial and harmful outcomes. One should see how the various alternatives have been compared against no treatment or placebo or among themselves. Some comparisons may be preferred or avoided and this may reflect biases. Moreover, instead of making 1 treatment comparison at a time, one may wish to analyze quantitatively all of the data from all comparisons together. If the trial results are compatible, the overall picture can help one to better appreciate the relative merits of all available interventions. This is very important for informing evidence-based guidelines and medical decision-making. Such a compilation of data from many systematic reviews and multiple meta-analyses is not something that can be performed lightly by a subject-matter expert based on subjective opinion alone. The synthesis of such complex information requires rigorous and systematic methods.

In this article, I review the main features, strengths and limitations of methods that integrate evidence across multiple meta-analyses. There are many new developments in systematic reviews in this area, but this article focuses on those that are becoming more influential in the literature: umbrella reviews and quantitative analyses of trial networks in which data are combined from clinical trials on diverse interventions for the same disease or condition. Readers are likely to see more of these designs published in medical journals and as background informing guidelines and recommendations. In a 1-year period (September 2007—September 2008), *CMAJ*, *The Lancet* and *BMJ* each published 1–2 network meta-analyses.^{5–8} Therefore, there is a need to understand the princi-

Key points

- Single meta-analysis of a treatment comparison for a single outcome offers a limited view if there are many treatments or many important outcomes to consider.
- Umbrella reviews assemble together several systematic reviews on the same condition.
- Treatment networks quantitatively analyze data for all treatment comparisons on the same disease.
- Multiple treatments meta-analysis can rank the effectiveness of many treatments in a network.
- Integration of evidence from multiple meta-analyses may be extended across many diseases.

ples behind these types of analysis. I also briefly discuss methods that analyze together data about several diseases and approaches that synthesize nonrandomized evidence from multiple meta-analyses.

Simple umbrella reviews

Umbrella reviews (Figure 1) are systematic reviews that consider many treatment comparisons for the management of the same disease or condition. Each comparison is considered separately, and meta-analyses are performed as deemed appropriate. Umbrella reviews are clusters that encompass many reviews. For example, an umbrella review presented data from 6 reviews that were considered to be of sufficiently high quality about nonpharmacological and nonsurgical interventions for hip osteoarthritis.⁹ Ideally, both benefits and harms should be juxtaposed to determine trade-offs between the risks and benefits.¹⁰

Few past reviews are explicitly called “umbrella reviews.” However, in the Cochrane collaboration, there is interest to assemble already existing reviews on the same topic under umbrella reviews. Moreover, many reviews already have features of umbrella reviews, even if not called umbrella reviews, if they consider many interventions and comparisons.

Compared with a systematic review or meta-analysis lim-

John Ioannidis is with the Clinical and Molecular Epidemiology Unit, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina, Greece; and the Institute for Clinical Research and Health Policy Studies, Department of Medicine, Tufts University School of Medicine, Boston, USA

Cite as *CMAJ* 2009. DOI:10.1503/cmaj.081086

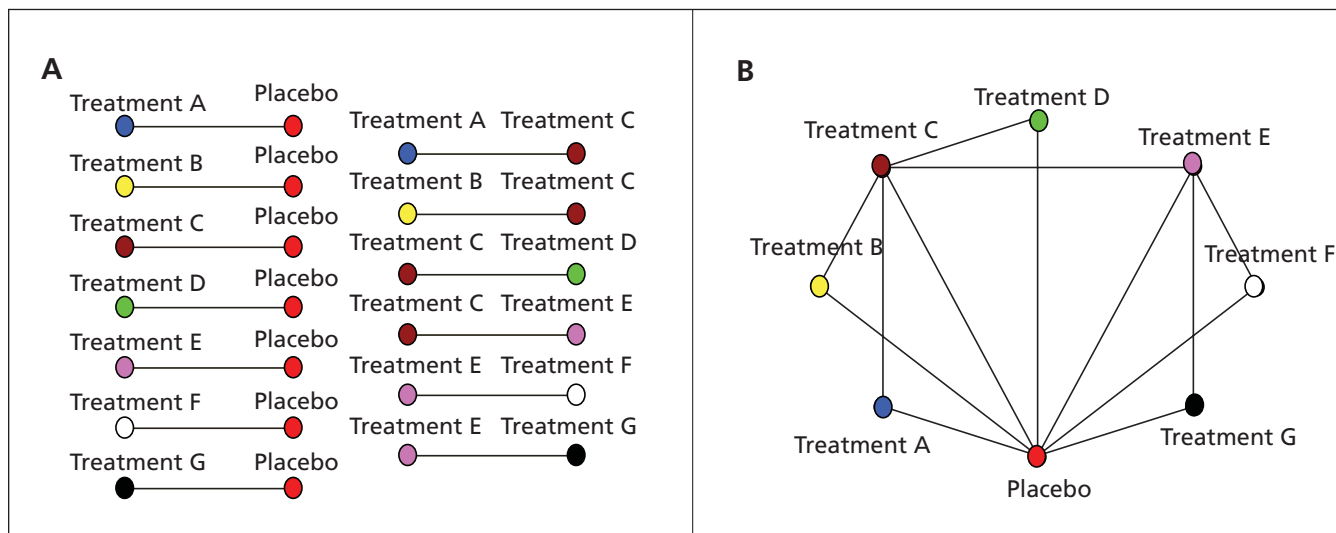


Figure 1: Schematic representations of (A) an umbrella review encompassing 13 comparisons involving 8 different treatment options (7 active treatments and placebo) and (B) a network with the same data. Each treatment is shown by a node of different colour, and comparisons between treatments are shown with links between the nodes. Each comparison may have data from several studies that may be combined in a traditional meta-analysis.

ited to 1 treatment comparison or even 1 outcome, an umbrella review can provide a wider picture on many treatments. This is probably more useful for health technology assessments that aim to inform guidelines and clinical practice where all the management options need to be considered and weighed. Conversely, some reviews may only address 1 drug, such as for narrow regulatory and licensing purposes, or even only 1 drug and 1 type of outcome for a highly focused question (e.g., whether rosiglitazone increases the risk of myocardial infarction).

Much like reviews on 1 treatment comparison, umbrella reviews are limited by the amount, quality and comprehensiveness of available information in the primary studies. In particular, a lack of data on harms^{11,12} and selective reporting of outcomes¹³⁻¹⁵ may lead to biased, optimistic pictures of the evidence. Patching together pre-existing reviews is limited by different eligibility criteria, evaluation methods and thoroughness of updating information across the merged reviews. Moreover, pre-existing reviews may not cover all of the possible management options. Finally, qualitative juxtaposition of data from separate meta-analyses is subjective and suboptimal. Umbrella reviews may be better to perform prospectively, defining upfront the range of interventions and outcomes that are to be addressed. This is a demanding but efficient investment of effort. The cumulative effort may be greater to perform each of the constituent reviews separately in a fragmented, uncoordinated fashion.

Treatment networks

A treatment network (Figure 1) shows together all of the treatment comparisons performed for a specific condition.^{16,17} It is depicted by the use of nodes for each available treatment and links between the nodes when the respective treatments

have been compared in 1 or more trials. A treatment network can be analyzed in terms of its geometry, and the outcome data from the included trials can be synthesized with multiple treatments meta-analysis.

Network geometry

Network geometry addresses what the shape of the network (nodes and links) looks like. If all of the treatments have been compared against placebo, but not among themselves, the network looks like a star with the placebo in its centre. If all of the treatment options have been compared with each other, the network is a polygon with all nodes connected to each other. There are many possible shapes between these extremes.

One can use measures that quantify network diversity and co-occurrence.¹⁶ Diversity is larger when there are many different treatments in the network and when the available treatments are more evenly represented (e.g., all have a similar amount of evidence). Consider the following example of limited diversity. Only 4 comparators (nicotine, bupropion, varenicline and placebo or no treatment) have been used across 174 compared arms of 84 drug trials of tobacco cessation. These options have been used very unevenly; there are 72, 14, 4 and 84 arms using these 4 options, respectively, in the available trials.¹⁸ The following is an example with more substantial diversity. Ten comparators have been used in 13 trials of topical antibiotics for chronic ear discharge, and none has been used in more than 8 trials.¹⁹ A simple index to measure diversity is the probability of interspecific encounter, which ranges between 0 and 1.¹⁶ In the first example with limited diversity, the probability of interspecific encounter is 0.59. The probability of interspecific encounter in the second example is 0.88.

Co-occurrence examines whether there is relative over- or under-representation of comparisons of specific pairs of treat-

Table 1: Examples of situations that may lead to limited diversity (few treatments or uneven representation of the available treatments) or significant co-occurrence in a treatment network

Situation	Limited diversity		Significant co-occurrence*
	Few treatments	Unevenness	
Few treatments available	+	–	–
Treatments available for evaluation for very different time periods	–	+	+
Sponsors pushing disproportionately specific treatments	–	+	–
Many treatments available, but evidence on some is suppressed (reporting bias)	+/-	+	+/-
Strong preference† to use a standard comparator	–	+	–
Strong preference for or avoidance of specific head-to-head comparisons	–	–	+
Many comparisons available, but evidence on some is suppressed (reporting bias)	–	+/-	+

*Co-occurrence means that some treatments are compared far more frequently against each other than expected based on their overall use in clinical trials in the network or that they are compared far less frequently against each other than expected. For example, if treatment A was used in 30 of the 60 trials in the network (50%) and treatment B was used in 18 of the 60 trials (30%), one would expect that $50\% \times 30\% = 15\%$ (9/60) trials should include a comparison of A and B, if the pairing of the treatments to be compared across trials was random. If all 18 trials of treatment B used A as the comparator, this is positive co-occurrence; if A was never compared against B, this is negative co-occurrence. Co-occurrence testing is based on permutations that consider not only 1 pair of treatments but all available treatments in the network. The available test may be underpowered to detect co-occurrence if few studies are available.¹⁶

†Preference may be justified (no active treatment shown to be effective so far), because of regulatory pressure (insistence on running placebo-controlled trials, even if effective treatments have been demonstrated previously) or unjustified (choosing a “straw man” comparator [placebo or inferior active treatment], avoidance of demonstrated effective comparators).

ments. In the smoking cessation example above,¹⁸ there were 69 trials that compared nicotine replacement with no active treatment. There was no comparison of varenicline and nicotine replacement and only 1 comparison of bupropion and nicotine replacement. Both varenicline and bupropion have been mostly compared with placebo, even though it is well documented that nicotine replacement is an effective treatment. A co-occurrence test (C test)¹⁶ for this network gives $p < 0.001$, which means that this distribution of comparisons is unlikely to be due to chance.

Many forces shape network geometry (Table 1). Sometimes very few treatments are available or the different treatments have been available for very different lengths of time; this may explain why there are different amounts of evidence and why the drugs have not been compared extensively against each other. Otherwise, limited diversity or significant co-occurrence often reflect “comparator preference biases” (i.e., some treatments or comparisons are particularly selected or avoided or information on them is suppressed [reporting biases]). A sponsor agenda may be revealed that may not necessarily be justified. There may be regulatory pressure to perform comparisons with placebo, or sponsors may try to avoid comparisons against other effective treatments and may prefer comparing their agents against inferior “straw men” (agents with poor effectiveness that are easy to beat). When examining network geometry, one should ask: Have the right types of comparisons been performed? Is there a treatment comparison for which there is little or no evidence but that would be essential to have evidence on?

Incoherence

Whenever a closed loop exists in a network, one may examine incoherence in the treatment comparisons that form this loop.^{20–22} Incoherence tells us whether the effect estimated from indirect comparisons differs from that estimated from

direct comparisons. The simplest loop involves 3 compared treatments. For example, before varenicline became available, only bupropion, nicotine replacement and placebo or no treatment were used in trials of smoking cessation. To estimate the effect of bupropion versus that of nicotine replacement, we

Box 1: Potential reasons for incoherence between the results of direct and indirect comparisons

- Chance
- Genuine diversity
 - Differences in enrolled participants (e.g., entry criteria, clinical setting, disease spectrum, baseline risk, selection based on prior response)
 - Differences in the exact interventions (e.g., dose, duration of administration, prior administration [second-line treatment])
 - Differences in background treatment and management (e.g., evolving treatment and management in more recent years)
 - Differences in exact definition or measurement of outcomes
- Bias in head-to-head comparisons
 - Optimism bias in favour of the new drug in appraising effectiveness
 - Publication bias
 - Selective reporting of outcomes and of analyses
 - Inflated effect size in early stopped trials and in early evidence
 - Defects in randomization, allocation concealment, masking or other key study design features
- Bias in indirect comparisons
 - Applicable to each of the comparisons involved in the indirect part of the loop

Box 2: Key features in the critical reading of umbrella reviews

- Are all pertinent interventions considered? It may not be practical to include every possible intervention, but are the most commonly used interventions and most important new interventions considered and is the choice rational and explicitly described?
- Are all pertinent outcomes (both beneficial and harmful) considered? These may include patient-centred outcomes and subjective outcomes as well as objective disease measures.
- What are the amount, quality and comprehensiveness of data? These might include patient-centred outcomes and subjective outcomes as well as objective disease measures. One must be cautious about the strength of any inferences made if there is limited data, if many of the included studies are of poor quality, if the evidence is fragmented with many studies not reporting on all important outcomes, or if there is concern for substantial publication bias.
- If pre-existing reviews are merged under the umbrella, do these reviews differ? The included reviews might differ in the populations covered, the time period, the types of studies summarized or the methods used in the reviews.
- Are the eligibility criteria described explicitly in a way that would enable a reader to repeat the selection?
- Are the evaluation methods described explicitly in a way that would enable a reader to repeat them?
- Thoroughness of updating of information: How and when was information updated and based on what sources and search strategies?
- How are data from separate meta-analyses compared (if not by formal multiple treatments meta-analysis)? Is the method explicit and described in a way that would enable a reader to repeat it, or is it based on subjective interpretation?

Box 3: Key considerations in the analysis of treatment networks and multiple treatments meta-analyses

- What is the network geometry? How have different treatments been compared against no treatment and against other treatments?
- Diversity: Are there many or few treatments being compared and is there a preference to use some specific treatments?
- Co-occurrence: Is there a tendency to avoid or to prefer comparisons between specific treatments?
- If there is limited diversity or significant co-occurrence, what are the possible explanations? Is limited diversity or significant co-occurrence justified or is it because of potential biases (e.g. comparator preference biases)?
- Is there incoherence in the network? Do the analyses of direct and indirect comparisons reach different conclusions? If so, what are the possible explanations?
- Is it reasonable to analyze all of the data together? Has the potential for clinical or other sources of heterogeneity and incoherence been properly considered?
- What are the treatment effects for each treatment comparison in the multiple treatments meta-analysis?
- What are the credibility intervals for the treatment effects (i.e., What is the uncertainty surrounding the estimate of each treatment effect?)
- What is the ranking of treatments (i.e., Which is the most likely treatment to be the best? How likely is it to be the best? How do the other treatment options rank?)
- Should the results be extrapolated to individual patients? Is the particular patient similar enough to the patients included in the meta-analysis? If not, are the dissimilarities important enough to question whether the results of the meta-analysis are meaningful to this particular patient?

can use either information from their direct (head-to-head) comparison or information from the comparison of each medication against placebo. When incoherence appears, one may ask why. Some reasons are listed in Box 1.

Direct randomized comparisons of treatments are usually more trustworthy than indirect comparisons,^{23,24} but not always.^{25–27} Red flags that may suggest that the direct comparisons are not necessarily as trustworthy as they should be include poor study design and the potential for conflicts of interest and optimism bias in favour of new drugs. For example, if conflicts of interest favour drug A over B, the direct comparison estimate is distorted, while indirect estimates of the effects of A versus those of B through comparator C may be unbiased. An empirical evaluation of head-to-head comparisons of antipsychotic drugs has shown that trials favour the drug of the sponsor 90% of the time.²⁷ Optimism bias in favour of new drugs may inflate their perceived effectiveness.²⁸ A biased research agenda may try to make a favoured drug “look nice.”^{29–31}

For example, 1 trial showed bupropion to be far better than nicotine patch for smoking cessation (odds ratio [OR] of smoking at 12 months 0.48, 95% confidence interval [CI] 0.28–0.82);²⁶ however, in trials of bupropion versus placebo and in trials of nicotine patch versus placebo, the indirect comparison of bupropion against nicotine patch suggested no major difference (OR 0.90, 95% CI 0.61–1.34).²⁶ The head-to-head comparison sponsored by the bupropion manufacturer might have overestimated its effectiveness.

A limitation is that the power to detect incoherence is low when there are only a few small trials. Moreover, results are still interpreted in a black or white fashion as “significant” or “nonsignificant” incoherence. Putting a number on what constitutes large incoherence is not yet feasible. Finally, explanation of incoherence has the limitations of any exploratory exercise.

Multiple treatments meta-analysis

Networks with closed loops can be analyzed by multiple treatments meta-analysis. Multiple treatments meta-analysis incorporates all of the data from both indirect and indirect comparisons of the treatments in the network. The computational details of this method are described elsewhere.^{17,20–22}

Most methods are Bayesian and use knowledge from the observed data to modify our understanding of how different treatments perform against each other and how treatments should be ranked. These methods are also able to explore inconsistencies in results between trials. A multiple treatments meta-analysis eventually derives treatment effects (e.g., relative risks) and the uncertainty (credibility intervals) for pair-wise comparisons of all of the treatments involved in the network, regardless of whether they were compared directly.

A useful feature is to rank the treatments on their probability of being the best. For example, a multiple treatments meta-analysis found that for systemic treatment of advanced colorectal cancer, combination of 5-fluorouracil, leucovorin, irinotecan and bevacizumab was 64% likely to be the most effective first-line regimen for extending survival and 95% likely to be 1 of the top 3 regimens (of 9 available).³²

Multiple treatments meta-analysis may include data that are equivalent to many traditional meta-analyses that address a single treatment comparison. For example, a multiple treatments meta-analysis on systemic treatment for advanced breast cancer³³ included data from 45 different direct comparisons, each of which could have been a separate traditional meta-analysis. Multiple treatments meta-analyses require more sophisticated statistical expertise than simple umbrella reviews, and data syntheses require caution in the presence of incoherence, but they allow objective, quantitative integration of the evidence. With diffusion of the proper expertise, these methods may partly replace (or upgrade) traditional meta-analyses.

A limitation of this method is that multiple treatments meta-analysis has to make a generous assumption that all of the data can be analyzed together (i.e., they are either similar enough or their dissimilarity can be properly taken into account during analysis). Some argue that incoherent data should not be combined, while others do not share this view. This is extension of the debate about whether heterogeneous data can be combined in traditional meta-analyses.³⁴ Finally, one has to critically consider whether the data can be extrapolated from large samples to individual patients and settings. With multiple treatments meta-analysis, evidence to inform on a specific treatment comparison is drawn even from entirely different treatment comparisons.

Box 2 and Box 3 list the key features of the critical reading and interpretation of umbrella reviews, trial networks and multiple treatments meta-analyses.

Other extensions

Examining together data about many conditions

Some reviews go a step further and consider not only diverse interventions on a given disease but also evidence on many diseases or conditions. These are called domain analyses³⁵ or meta-epidemiologic research.³⁶⁻³⁸ Such analyses can offer hints about the reliability of treatment effects in whole fields; for example, whether treatment effects are reliable,^{39,40} whether they change over time⁴¹ and whether they are related to some study characteristics, regardless of the disease.^{38,42-46}

Nonrandomized evidence

Some of the concepts discussed here have parallel applications to nonrandomized research, including diagnostic, prognostic and epidemiologic associations. For example, field synopses are the equivalent of umbrella reviews for epidemiologic associations. Field synopses perform systematic reviews and meta-analyses on all associations in a field. For example, AlzGene⁴⁷ and SzGene⁴⁸ are field synopses on genetic associations for Alzheimer disease and schizophrenia, respectively. Data from over 1000 studies and on over 100 associations are summarized in each synopsis. Given that, for many diseases, there can be hundreds of postulated associations (e.g., genetic, nutritional, environmental), systematizing this knowledge is essential to keep track of where we stand and what to make of the torrents of data on postulated risk factors.

In theory, umbrella reviews may also encompass reviews and meta-analyses on data of diagnostic, prognostic and predictive tests, if these are pertinent to consider in the overall management of a disease, in addition to just treatment decisions. Also, networks and multiple treatments meta-analysis may encompass nonrandomized data, but this is rarely considered, perhaps because these data are deemed different from those that result from randomized trials. Finally, domain analyses can be useful in understanding biases in observational research.^{49,50} There are also examples of meta-epidemiologic research in the prognostic⁵¹ and diagnostic^{52,53} literature.

Conclusion

Integrating data from multiple meta-analyses may provide a wide view of the evidence landscape. Transition from a single patient to a study of many patients is a leap of faith in generalizability. A further leap is needed for the transition from a single study to meta-analysis and from a traditional meta-analysis to a treatment network and multiple treatments meta-analysis, let alone wider domains. With this caveat, zooming out toward larger scales of evidence may help us to understand the strengths and limitations of the data guiding the medical care of individual patients.

This article has been peer reviewed.

Competing interests: None declared.

REFERENCES

1. Patsopoulos NA, Analatos AA, Ioannidis JP. Relative citation impact of various study designs in the health sciences. *JAMA* 2005;293:2362-6.
2. Anatomical Therapeutic Chemical (ATC) classification. Oslo, Norway: WHO Collaborating Centre for Drug Statistics Methodology and Norwegian Institute of Public Health; 2008. Available: www.whocc.no/atcddd/ (accessed 2009 June 8).
3. Ioannidis JP. Effectiveness of antidepressants: an evidence myth constructed from a thousand randomized trials? *Philos Ethics Humanit Med* 2008;3:14.
4. Kirsch I, Deacon BJ, Huedo-Medina TB, et al. Initial severity and antidepressant benefits: a meta-analysis of data submitted to the Food and Drug Administration. *PLoS Med* 2008;5:e45.
5. Stettler C, Allemann S, Wandel S, et al. Drug eluting and bare metal stents in people with and without diabetes: collaborative network meta-analysis. *BMJ* 2008;337:a1331.
6. Eisenberg MJ, Filion KB, Yavin D, et al. Pharmacotherapies for smoking cessation: a meta-analysis of randomized controlled trials. *CMAJ* 2008;179:135-44.
7. Lam SK, Owen A. Combined resynchronisation and implantable defibrillator therapy in left ventricular dysfunction: Bayesian network meta-analysis of randomised controlled trials. *BMJ* 2007;335:925.
8. Stettler C, Wandel S, Allemann S, et al. Outcomes associated with drug-eluting

- and bare-metal stents: a collaborative network meta-analysis. *Lancet* 2007;370:937-48.
9. Moe RH, Haavardsholm EA, Christie A, et al. Effectiveness of nonpharmacological and nonsurgical interventions for hip osteoarthritis: an umbrella review of high-quality systematic reviews. *Phys Ther* 2007;87:1716-27.
 10. Glasziou PP, Irwig LM. An evidence based approach to individualising treatment. *BMJ* 1995;311:1356-9.
 11. Ioannidis JP, Lau J. Completeness of safety reporting in randomized trials: an evaluation of 7 medical areas. *JAMA* 2001;285:437-43.
 12. Papanikolaou PN, Ioannidis JP. Availability of large-scale evidence on specific harms from systematic reviews of randomized trials. *Am J Med* 2004;117:582-9.
 13. Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ* 2005;330:753.
 14. Chan AW, Krieza-Jeri K, Schmid I, et al. Outcome reporting bias in randomized trials funded by the Canadian Institutes of Health Research. *CMAJ* 2004;171:735-40.
 15. Chan AW, Hróbjartsson A, Haahr MT, et al. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65.
 16. Salanti G, Kavvoura FK, Ioannidis JP. Exploring the geometry of treatment networks. *Ann Intern Med* 2008;148:544-53.
 17. Salanti G, Higgins JP, Ades A, et al. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008;17:279-301.
 18. Wu P, Wilson K, Dimoulas P, et al. Effectiveness of smoking cessation therapies: a systematic review and meta-analysis. *BMC Public Health* 2006;6:300.
 19. Macfadyen CA, Acuin JM, Gamble C. Topical antibiotics without steroids for chronically discharging ears with underlying eardrum perforations. *Cochrane Database Syst Rev* 2005;(4):CD004618.
 20. Lu G, Ades AE. Assessing evidence consistency in mixed treatment comparisons. *J Am Stat Assoc* 2006;101:447-59.
 21. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med* 2002;21:2313-24.
 22. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005;331:897-900.
 23. Song F, Altman DG, Glenny AM, et al. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ* 2003;326:472.
 24. Glenny AM, Altman DG, Song F, et al.; International Stroke Trial Collaborative Group. Indirect comparisons of competing interventions. *Health Technol Assess* 2005;9:1-134, iii-iv.
 25. Ioannidis JP. Indirect comparisons: the mesh and mess of clinical trials. *Lancet* 2006;368:1470-2.
 26. 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;61:455-63.
 27. Heres S, Davis J, Maino K, et al. Why olanzapine beats risperidone, risperidone beats quetiapine, and quetiapine beats olanzapine: an exploratory analysis of head-to-head comparison studies of second-generation antipsychotics. *Am J Psychiatry* 2006;163:185-94.
 28. Chalmers I, Matthews R. What are the implications of optimism bias in clinical research? *Lancet* 2006;367:449-50.
 29. Bero L, Oostvogel F, Bacchetti P, et al. Factors associated with findings of published trials of drug-drug comparisons: why some statins appear more efficacious than others. *PLoS Med* 2007;4:e184.
 30. Peppercorn J, Blood E, Winer E, et al. Association between pharmaceutical involvement and outcomes in breast cancer clinical trials. *Cancer* 2007;109:1239-46.
 31. Ioannidis JP. Perfect study, poor evidence: interpretation of biases preceding study design. *Semin Hematol* 2008;45:160-6.
 32. Golfinopoulos V, Salanti G, Pavlidis N, et al. Survival and disease-progression benefits with treatment regimens for advanced colorectal cancer: a meta-analysis. *Lancet Oncol* 2007;8:898-911.
 33. Mauri D, Polyzos NP, Salanti G, et al. Multiple treatments meta-analysis of chemotherapy and targeted therapies in advanced breast cancer. *J Natl Cancer Inst* 2008;100:1780-91.
 34. Ioannidis JP, Patsopoulos N, Rothstein H. Reasons or excuses for avoiding meta-analysis in forest plots. *BMJ* 2008;336:1413-5.
 35. Ioannidis JP, Trikalinos TA. An exploratory test for an excess of significant findings. *Clin Trials* 2007;4:245-53.
 36. Siersma V, Als-Nielsen B, Chen W, et al. Multivariable modelling for meta-epidemiological assessment of the association between trial quality and treatment effects estimated in randomized clinical trials. *Stat Med* 2007;26:2745-58.
 37. Sterne JA, Juni P, Schulz KF, et al. Statistical methods for assessing the influence of study characteristics on treatment effects in 'meta-epidemiological' research. *Stat Med* 2002;21:1513-24.
 38. Wood L, Egger M, Gluud LL, 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.
 39. Shang A, Huwiler-Müntener K, Nartey L, et al. Are the clinical effects of homeopathy placebo effects? Comparative study of placebo-controlled trials of homeopathy and allopathy. *Lancet* 2005;366:726-32.
 40. Ioannidis JP. Why most discovered true associations are inflated. *Epidemiology* 2008;19:640-8.
 41. Ioannidis J, Lau J. Evolution of treatment effects over time: empirical insight from recursive cumulative metaanalyses. *Proc Natl Acad Sci U S A* 2001;98:831-6.
 42. Schulz KF, Chalmers I, Hayes RJ, et al. Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* 1995;273:408-12.
 43. Balk EM, Bonis PA, Moskowitz H, et al. Correlation of quality measures with estimates of treatment effect in meta-analyses of randomized controlled trials. *JAMA* 2002;287:2973-82.
 44. Moher D, Pham B, Jones A, et al. Does quality of reports of randomised trials affect estimates of intervention efficacy reported in meta-analyses? *Lancet* 1998;352:609-13.
 45. 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.
 46. Juni P, Altman DG, Egger M. Systematic reviews in health care: Assessing the quality of controlled clinical trials. *BMJ* 2001;323:42-6.
 47. Bertram L, McQueen MB, Mullin K, et al. Systematic meta-analyses of Alzheimer disease genetic association studies: the AlzGene database. *Nat Genet* 2007;39:17-23.
 48. Allen NC, Bagade S, McQueen MB, et al. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. *Nat Genet* 2008;40:827-34.
 49. Kyzas PA, Denaxa-Kyza D, Ioannidis JP. Almost all articles on cancer prognostic markers report statistically significant results. *Eur J Cancer* 2007;43:2559-79.
 50. Kavvoura FK, McQueen M, Khoury MJ, et al. Evaluation of the potential excess of statistically significant findings in reported genetic association studies: application to Alzheimer's disease. *Am J Epidemiol* 2008;168:855-65.
 51. Kyzas PA, Denaxa-Kyza D, Ioannidis JP. Quality of reporting of cancer prognostic marker studies: association with reported prognostic effect. *J Natl Cancer Inst* 2007;99:236-43.
 52. Lijmer JG, Mol BW, Heisterkamp S, et al. Empirical evidence of design-related bias in studies of diagnostic tests. *JAMA* 1999;282:1061-6.
 53. Rutjes AW, Reitsma JB, Di Nisio M, et al. Evidence of bias and variation in diagnostic accuracy studies. *CMAJ* 2006;174:469-76.

Correspondence to: Dr. John P.A. Ioannidis, Professor and Chairman, Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, Ioannina 45110, Greece; fax 30 26510 97867; jioannid@cc.uoi.gr