REVIEW ARTICLE

Indirect Comparisons and Network Meta-Analyses

Estimation of Effects in the Absence of Head-to-Head Trials—Part 22 of a Series on Evaluation of Scientific Publications

Corinna Kiefer, Sibylle Sturtz, Ralf Bender

SUMMARY

Background: Systematic reviews provide a structured summary of the results of trials that have been carried out on any particular subject. If the data from multiple trials are sufficiently homogenous, a meta-analysis can be performed to calculate pooled effect estimates. Traditional meta-analysis involves groups of trials that compare the same two interventions directly (head to head). Lately, however, indirect comparisons and network meta-analyses have become increasingly common.

Methods: Various methods of indirect comparison and network meta-analysis are presented and discussed on the basis of a selective review of the literature. The main assumptions and requirements of these methods are described, and a checklist is provided as an aid to the evaluation of published indirect comparisons and network meta-analyses.

Results: When no head-to-head trials of two interventions are available, indirect comparisons and network meta-analyses enable the estimation of effects as well as the simultaneous analysis of networks involving more than two interventions. Network meta-analyses and indirect comparisons can only be useful if the trial or patient characteristics are similar and the observed effects are sufficiently homogeneous. Moreover, there should be no major discrepancy between the direct and indirect evidence. If trials are available that compare each of two treatments against a third one, but not against each other, then the third intervention can be used as a common comparator to enable a comparison of the other two.

Conclusion: Indirect comparisons and network meta-analyses are an important further development of traditional meta-analysis. Clear and detailed documentation is needed so that findings obtained by these new methods can be reliably judged.

Medical Biometry Department, Institute for Quality and Efficiency in Health Care (IQWiG), Köln: Kiefer, Dr. rer. nat. Sturtz, Prof. Dr. rer. biol. hum. Bender

Reviews are often used in medical research to collate, evaluate, and summarize the evidence on a particular clinical question systematically and transparently (1). To compare exactly two interventions, the results of available head-to-head trials (often randomized controlled trials) are quantitatively summarized in a meta-analysis (1–3). However, this approach becomes problematic when there are no head-to-head trials comparing the two interventions, or when more than two interventions need to be compared with each other simultaneously. For comparison of newer anticoagulants in patients with atrial fibrillation, for example, although there are trials directly comparing each of the newer drugs with the current standard treatment (warfarin), there are none that compare two of the newer anticoagulants with each other directly (4). Another example is the research of various prostaglandins to induce labor. Here, 14 interventions can be compared with each other. Rather than using numerous separate pairwise comparisons, it is preferable to perform a combined data analysis (5). Cases such as these require procedures for indirect comparison or network meta-analysis.

Since 2009, indirect comparisons and network analyses have become increasingly important (6). In addition to the well-known PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) statement (7, 8), a guideline has also been compiled for the publication of systematic reviews incorporating network meta-analyses (9).

This article aims to describe the underlying assumptions and methods used in indirect comparisons and network meta-analyses and to explain what evaluation of such publications should include.

Methods

Below is an explanation of the various terms, essential statistical procedures, and assumptions made for indirect comparisons and network meta-analyses. A selective search of the literature was performed for this purpose.

Terms

As yet there is no single term in the literature for the various methods used to perform indirect comparisons (10). In this article, as in (11), the expression "methods

Adjusted indirect comparison according to Bucher et al. (13)

The effect of intervention B relative to intervention A can be estimated indirectly as follows, using the direct estimators for the effects of intervention C relative to intervention A ($effect_{AC}$) and intervention C relative to intervention B ($effect_{BC}$):

$$effect_{AB} = effect_{AC} - effect_{BC}$$

The variance of the indirect estimator $effect_{AB}$ is the sum of the variances of the direct estimators:

$$variance_{AB} = variance_{AC} + variance_{BC}$$

The corresponding two-tailed 95% confidence interval can thus be calculated as follows:

$$\left[effect_{AB} - Z_{0.975} \times \sqrt{variance_{AB}};\ effect_{AB} + Z_{0.975} \times \sqrt{variance_{AB}} \right]$$

$Z_{0.975}$ here refers to the 97.5% quantile of standard normal distribution, which gives a rounded value of 1.96.

for indirect comparison" in the broader sense includes both procedures used for simple indirect comparisons of two interventions and those used to compare more than two interventions and/or to combine direct and indirect evidence. For the latter case, the terms "mixed treatment comparison meta-analysis," "multiple treatment meta-analysis," and "network meta-analysis" are used in the literature (10). It seems sensible to use the term "network meta-analysis" when more than two interventions are to be compared with each other. Indirect comparison in the narrower sense is the synthesis of purely indirect evidence concerning two interventions.

Statistical procedures

The general scientific consensus is that it is inappropriate to use nonadjusted indirect comparisons in which findings from individual arms of different trials are naively compared with each other without taking randomization into account (11, 12). This article therefore describes only procedures for adjusted indirect comparisons which further analyze, for example, the effects estimated in the trials.

If there is no evidence available from head-to-head trials comparing two interventions, but trials comparing each of the interventions of interest with the same comparator (e.g. a placebo) have been conducted, the method of adjusted indirect comparison as described by Bucher et al. (13) is appropriate.

Part (a) of the *Figure* shows the simplest case of a network with three interventions, A, B, and C. The lines represent head-to-head trials. The findings of the head-to-head trial of A and C, and of a second head-to-head trial of B and C, can be used to compare interventions A

and B with each other indirectly. Intervention C is therefore the common comparator of the two interventions of interest (the common comparator). The effect of intervention B relative to A can be estimated indirectly for absolute effect measures (e.g. differences from the means, differences in risk) using the method put forward by Bucher et al. (13), as shown in *Box 1*. For relative effect measures (e.g. odds ratio, relative risk), it has to be taken into account that this additive relation holds true only on a logarithmic scale. The variance of the indirect estimator is the sum of the variances of the two direct estimators. The method described by Bucher et al. (13) can also be used for star-shaped networks (part [b] of the *Figure*) if only two-armed trials are available. If more than one trial comparing two particular interventions is available, the trials are first synthesized in a meta-analysis, and the corresponding estimator and its variance are used.

For more complex networks (part [c] of the *Figure*), the method put forward by Bucher et al. (13) becomes increasingly difficult and ultimately impossible to use (10), as it is unsuitable for combining evidence from direct and indirect comparisons or for multi-armed trials. In these cases, more complex models must be used in the framework of network meta-analysis. Network meta-analyses provide effect estimates for all possible pairwise comparisons within the network. To do this, the available direct and indirect evidence is combined simultaneously for every pairwise analysis. Data analysis can be performed using either a frequentist or a Bayesian approach (14). Various aspects can be particularly important, depending on the choice of procedure. If Bayesian methods are used, the choice of prior information in the form of a priori distributions plays a particularly vital role.

Basic assumptions

Underlying all indirect comparisons are three basic assumptions. The first two also apply to pairwise meta-analysis.

- Firstly, all the trials included must be comparable in terms of potential effect modifiers (e.g. trial or patient characteristics) (assumption of similarity).
- Secondly, there must be no relevant heterogeneity between trial results in pairwise comparisons (assumption of homogeneity).
- Thirdly, there must be no relevant discrepancy or inconsistency between direct and indirect evidence (assumption of consistency).

Assessment of published indirect comparisons and network meta-analyses

The essential issues to bear in mind when evaluating published indirect comparisons and network meta-analyses are described below. In part, these necessarily overlap with the criteria for evaluating systematic reviews containing traditional meta-analyses (3). The resulting checklist *(Box 2)* allows the reader to evaluate whether a published indirect comparison or network meta-analysis meets certain criteria. This article does not include technical details, particularly on statistical

FIGURE

Examples of network diagrams
a) Simple indirect comparison
b) Star-shaped network
c) More complex network containing 6 interventions
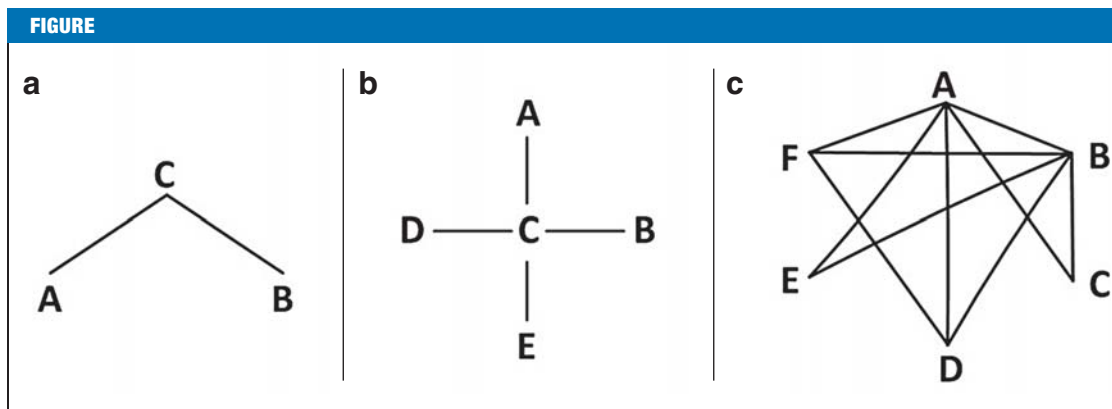
methods. The interested reader is referred to the available literature (10, 14–18).

1. Is the question to be addressed established in advance?
As is customary in systematic reviews, the underlying question addressed, and its expression in the form of statistical hypotheses should be clearly defined and established in advance, in writing, in a study protocol. Deviations from the procedure originally planned must be suitably described and justified.

2. Is sufficient rationale given for the use of indirect comparisons?
Indirect comparisons usually allow less certain conclusions compared to syntheses of direct evidence (11). There should thus always be sufficient explanation of the reasons for using indirect comparisons to address a clinical question.

3. Is sufficient rationale given for the choice of common comparators?
The outcomes of indirect comparisons depend to a great extent on the choice of common comparators, among other factors. If there are several options available, it may be possible to steer findings deliberately in a desirable direction. The choice of common comparators must therefore be suitably explained.

4. Has a complete, systematic search of the literature been performed and described in detail?
As with systematic reviews that include conventional meta-analyses, a complete, systematic search of the literature must be performed (7, 8). The relevant literature must be complete not only with regard to the interventions of interest but also with regard to the common comparators.

5. Have pre-established trial inclusion and exclusion criteria been used, and have they been clearly described?
As is customary in systematic reviews, inclusion and exclusion criteria are used to decide which of the trials identified during the search of the literature are to be included. Ideally, identification of the relevant literature should be clearly represented in the form of a flowchart (7, 8).

6. Was a complete report of all the relevant data available?
To avoid biased results, a complete report is needed. This entails a full representation of the essential trial characteristics, trial evaluations, individual trial results (estimates of effect and confidence intervals), and pooled effect estimates including confidence intervals for all relevant outcomes, comparisons, and subgroups. For more complex networks at least, a diagram of the network should also be included, with a description of the geometry of the network, i.e. its essential properties (9, 19, 20).

7. Have the basic assumptions of similarity, homogeneity, and consistency been examined, and have the findings of this examination been suitably handled?
As with pairwise meta-analyses, the similarity of the individual trials included should be examined on the basis of their essential characteristics; however, in this case this must be done for all the investigated interventions. The well-known PICOS approach—population (P), intervention (I), comparator (C), outcome (O), study design (S)—plays a vital role in this (7, 8). Important information can be obtained from comparisons of trials regarding relevant patient characteristics as well as comparisons of trial arms representing reasonable reference interventions regarding relevant endpoints.

As with pairwise meta-analyses, homogeneity should be examined using standard procedures such as forest plots and measures of heterogeneity, with defining criteria established in advance (7, 8). Depending on the size of the network, however, this may be a very lengthy process, as all possible combinations of two interventions must be included.

Consistency can usually only be examined if direct and indirect evidence is available for comparison

Checklist for evaluation of indirect comparisons and network meta-analyses

1. Were the questions to be addressed established in advance?
   – Clear description of question to be addressed
   – Expression in form of statistical hypotheses
   – Justification of deviations from procedure originally planned

2. Is sufficient rationale given for the use of indirect comparisons?

3. Is sufficient rationale given for the choice of common comparators?

4. Has a complete, systematic search of the literature been performed and described in detail?
   – For the interventions of primary interest?
   – For the common comparators?

5. Have pre-established trial inclusion and exclusion criteria been used, and have they been clearly described?

6. Was a complete report of all the relevant data available?
   – Characteristics of all included trials
   – Evaluations of all included trials
   – Diagram of network, description of network geometry
   – For all relevant outcomes, comparisons, and subgroups:
     – Results of all individual trials (effect estimates and confidence intervals)
     – Effect estimates and confidence intervals from pairwise meta-analyses

7. Have the basic assumptions been examined, and have the findings of this examination been suitably handled?
   – Similarity
   – Homogeneity
   – Consistency

8. Have suitable statistical procedures been used and described in detail?
   – Use of adjusted indirect comparisons
   – Handling of multi-armed studies
   – Random-effects or fixed-effect models
   – Technical details (particularly for Bayesian approaches)
   – Program code
   – Sensitivity analyses

9. Have limitations been sufficiently described and discussed?
   – Data quality and completeness
   – Methodological uncertainties, sensitivity analyses
   – Violations of basic assumptions

simultaneously. This means that simple indirect comparisons involving only one common comparator always entail greater uncertainty, as one of the three basic assumptions cannot be examined. There is a range of procedures for examining consistency (16, 21–24). The support of an experienced biometrician is required to assess whether these procedures have been used appropriately.

The results of examination of similarity, homogeneity, and consistency must be handled appropriately. If there has been a major violation of any of these assumptions, this may even mean that no worthwhile indirect comparison can be performed (23). In some cases, a change to the network, its division into subgroups, or the use of relevant co-variables may make a worthwhile indirect comparison possible (10, 25).

8. Have suitable statistical procedures been used and described in detail?

Because some procedures for network meta-analysis are complex, they are particularly challenging to evaluate. Such evaluation can only be performed by an experienced biometrician. This requires a highly detailed description of the methods used (9) and often also a representation of program code. Published checklists provide help for this (26–28). Important aspects are as follows:
- Whether adjusted indirect comparisons have been used
- Whether appropriate fixed-effect or random-effects models have been selected
- Whether multi-armed trials have been handled appropriately
- Whether unclear issues have been examined using sensitivity analyses.

Depending on whether a Bayesian or a frequentist approach has been used, further technical details that cannot be described in more detail here may play a role.

9. Have limitations been sufficiently described and discussed?

As in any scientific research, potential limitations that limit the certainty of the findings should be sufficiently reported and discussed. For indirect comparisons and network meta-analyses, data quality and completeness, methodical uncertainties and sensitivity analyses associated with these, and any violations of the three basic assumptions are particularly important. These should be evaluated by an experienced biometrician.

Results

The described checklist *(Box 2)* has been applied to published adjusted indirect comparisons and a published network meta-analysis as examples. The findings of this are shown in the *Table*. Baker and Phung (4) investigated newer anticoagulants in patients with atrial fibrillation using adjusted indirect comparisons. They included a total of four randomized controlled trials comparing apixaban (one study), dabigatran (two studies), and rivaroxaban (one study) with the current standard treatment, warfarin. Warfarin was selected as the common comparator for the indirect comparisons. Alfirevic et al. (5) performed a Bayesian network meta-analysis to compare the benefit and harm of different various prostaglandins to induce labor. Depending on the available data for each outcome, networks with up to 14 interventions and a dataset consisting of a total of 280 trials were evaluated.

Both publications are examples of essentially thorough conduct and reporting of indirect comparisons. However, in places more detail would have been desirable. The use of the checklist *(Box 2)* also revealed the following limitations.

Baker and Phung (4) did not provide sufficient justification for the choice of warfarin as common comparator. Although they did state that warfarin was the current standard treatment for which at least non-inferiority had already been shown for all three of the drugs included in the comparisons, it is unclear whether this choice was selective or whether warfarin is actually the only substance for which head-to-head trials of all three of the drugs of interest were available. A placebo could also have been used as a common comparator, but this was not discussed further by the authors.

Both publications describe a detailed information-gathering process. However, Baker and Phung (4) have not provided a list of the excluded references. Alfirevic et al. (5) refer to only one secondary source (the Cochrane Pregnancy and Childbirth Group Register), the update status of which is unclear. There is no clear representation of how this secondary source was searched. It therefore remains uncertain whether the study pool is up to date and complete.

Both publications lack an examination of the assumption of homogeneity in their pairwise meta-analyses. Alfirevic et al. (5) examined homogeneity only as part of network meta-analysis for all pairwise comparisons combined; however, this is not sufficient. Baker and Phung (4) also fail to provide a representation of the pooled results of the comparison of dabigatran and warfarin. Because the assumption of homogeneity has not been examined, it is possible that pairwise comparisons were associated with substantial heterogeneity between trial findings in both publications.

Discussion

Indirect comparisons and network meta-analyses make it possible to estimate effects in order to compare different interventions in systematic reviews even if there are no head-to-head trials of them. They also allow for simultaneous analysis of networks containing more than two interventions. Adjusted indirect comparisons can be used, for example, to compare the benefits and harms of newer anticoagulants (apixaban, dabigatran, and rivaroxaban) with each other in patients with atrial fibrillation even though there are no head-to-head trials comparing them. Network meta-analysis also makes it possible to compare simultaneously, in a joint evaluation, the benefits and harms of 14 interventions (prostaglandins, no treatment, placebo) to induce labor. Indirect comparisons and network meta-analyses can thus provide findings of fundamental importance for the development of guidelines and for evidence-based decisions in health care. They therefore represent important extensions to traditional pairwise meta-analyses. However, appropriate use of these methods requires strict assumptions that often cannot be fully verified on the basis of the available data. The findings of indirect comparisons and network meta-analyses therefore usually allow for less certainty in conclusions than the findings of appropriate pairwise meta-analyses of head-to-head trials. Transparent, detailed documen-

tation is required so that the published findings of indirect comparisons and network meta-analyses can be suitably evaluated. The simple checklist presented in this article provides an aid for this.

TABLE

Evaluation of indirect comparisons and network meta-analyses in data examples

Checklist item	Baker and Phung (4)	Alfivirec et al. (5)
1. Questions established in advance?	✓	✓
2. Rationale given for use of indirect comparisons?	✓	✓
3. Rationale given for choice of common comparators?	— [*1]	✓
4. Search of literature complete?	(✓) [*2]	(✓) [*5]
5. Inclusion and exclusion criteria clearly described?	✓	✓
6. Complete report of all data?	(✓) [*3]	✓
7. Basic assumptions examined?	— [*4]	— [*4]
8. Suitable statistical procedures used?	✓	✓
9. Limitations described?	✓	✓

[*1] It is unclear whether or not this is a selective choice.
[*2] There is no list of excluded references.
[*3] There is no representation of pooled estimates for pairwise comparison of dabigatran and warfarin.
[*4] Homogeneity has not been sufficiently examined.
[*5] The method used to search the secondary source is not clearly shown. It is therefore unclear whether the study pool up is up to date and complete.

REFERENCES

1. Egger M, Davey Smith G, Altman DG (eds.): Systematic reviews in health care: meta-analysis in context. London: BMJ Books 2001.

2. Ziegler A, Lange S, Bender R: Systematische Übersichten und Meta-Analysen. Dtsch Med Wochenschr 2007; 132 (Suppl. 1): e48–52.

3. Ressing M, Blettner M, Klug SJ: Systematic literature reviews and meta-analyses: Part 6 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009; 106: 456–63.

4. Baker WL, Phung OJ: Systematic review and adjusted indirect comparison meta-analysis of oral anticoagulants in atrial fibrillation. Circ Cardiovasc Qual Outcomes 2012; 5: 711–9.

5. Alfirevic Z, Keeney E, Dowswell T, et al.: Labour induction with prostaglandins: A systematic review and network meta-analysis. BMJ 2015; 350: h217.

6. Lee AW: Review of mixed treatment comparisons in published systematic reviews shows marked increase since 2009. J Clin Epidemiol 2014; 67: 138–43.

KEY MESSAGES

- Indirect comparisons and network meta-analyses make it possible to estimate effects in systematic reviews when there are no head-to-head trials or when multiple interventions are to be compared with each other simultaneously.

- The basic assumptions for indirect comparisons and network meta-analyses are sufficient similarity, homogeneity, and consistency.

- Nonadjusted indirect comparisons, i.e. naive comparisons of individual arms of different trials, are not an appropriate method of analysis.

- Transparent, detailed documentation is required so that the published findings of indirect comparisons and network meta-analyses can be suitably evaluated.

- An experienced biometrician is required for evaluation of statistical methods used for indirect comparisons and network meta-analyses.

7. Moher D, Liberati A, Tetzlaff J, Altman DG: Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. BMJ 2009; 339: b2535.

8. Liberati A, Altman DG, Tetzlaff J, et al.: The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: explanation and elaboration. BMJ 2009; 339: b2700.

9. Hutton B, Salanti G, Caldwell DM, et al.: The PRISMA extension statement for reporting of systematic reviews incorporating network meta-analyses of health care interventions: Checklist and explanations. Ann Intern Med 2015; 162: 777–84.

10. 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 Syn Methods 2012; 3: 80–97.

11. Bender R, Schwenke C, Schmoor C, Hauschke D: Stellenwert von Ergebnissen aus indirekten Vergleichen – Gemeinsame Stellungnahme von IQWiG, GMDS und IBS-DR. Köln, 02. Februar 2012. www.gmds.de/pdf/publikationen/stellungnahmen/120202_IQWIG_GMDS_IBS_DR.pdf (last accessed on 21 September 2015).

12. Higgins JPT, Deeks JJ, Altman DG, on behalf of the Cochrane Statistical Methods Group (eds.): Special topics in statistics. In: Higgins JPT, Deeks JJ, Altman DG, Green S (eds.): Cochrane Handbook for Systematic Reviews of Interventions. Chichester: Wiley 2008; 481–529.

13. 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.

14. Hoaglin DC, Hawkins N, Jansen JP, et al.: Conducting indirect-treatment-comparison and network-meta-analysis studies: Report of the ISPOR Task Force on Indirect Treatment Comparisons Good Research Practices: Part 2. Value Health 2011; 14: 429–37

15. 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.

16. 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: 641–56.

17. Rücker G: Network meta-analysis, electrical networks and graph theory. Res Syn Methods 2012; 3: 312–24.

18. Salanti G, Higgins JPT, Ades A, Ioannidis JPA: Evaluation of networks of randomized trials. Stat Methods Med Res 2007; 17: 279–30.

19. Salanti G, Kavvoura FK, Ioannidis JPA: Exploring the geometry of treatment networks. Ann Intern Med 2008; 148: 544–53.

20. Rücker G, Schwarzer G: Automated drawing of network plots in network meta-analysis. Res Syn Methods 2015 Jun 9. DOI: 10.1002/jrsm.1143 (Epub ahead of print).

21. Dias S, Welton NJ, Caldwell DM, Ades AE: Checking consistency in mixed treatment comparison meta-analysis. Stat Med 2010; 29: 932–44.

22. 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 Syn Methods 2012; 3: 98–110.

23. Donegan S, Williamson P, D'Alessandro U, Tudur Smith C: Assessing key assumptions of network meta-analysis: A review of methods. Res Syn Methods 2013; 4: 291–323.

24. Krahn U, Binder H, König J: A graphical tool for locating inconsistency in network meta-analyses. BMC Med Res Methodol 2013; 13: 35.

25. Sturtz S, Bender R: Unsolved issues of mixed treatment comparison meta-analysis: Network size and inconsistency. Res Syn Methods 2012; 3: 300–11.

26. Jansen JP, Fleurence R, Devine B, 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.

27. Jansen JP, Trikalinos T, Cappelleri JC, 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.

28. 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.

Corresponding author:
Prof. Dr. rer. biol. hum. Ralf Bender
Institute for Quality and Efficiency in Health Care (IQWiG)
Im Mediapark 8
50670 Köln, Germany
Ralf.Bender@iqwig.de