CrossMark

EDITORIAL

Network meta-analyses should be the highest level of evidence in treatment guidelines

Stefan Leucht[1] · Anna Chaimani[2] · Andrea S. Cipriani[3] · John M. Davis[4,5] ·
Toshi A. Furukawa[6] · Georgia Salanti[7]

After initial hesitancy due to fears that this procedure might lead to "cookbook medicine" and others, evidence-based medicine (EBM) is now an accepted principle in all fields of medicine including psychiatry. The essence of the evidence is used by many treatment guidelines to inform clinicians in their daily practice. One not entirely resolved issue is, however, which study or evidence synthesis design should be considered as the highest level of evidence. Early statements from McMaster University in Canada [5] (together with the Cochrane Collaboration, the "cradle" of EBM) suggested systematic reviews with meta-analysis can provide the most robust and reliable evidence, but not all guideline producers are in agreement. This is a timely debate, fuelled by the increasing publication of network meta-analyses, a novel approach which takes the assumptions of meta-analysis one step further [3]. Conventional meta-analyses only average the randomised trials comparing two treatments directly (so-called direct evidence). The major criticism has been that meta-analysis compares "apples and oranges"; are trials sufficiently similar so that they can be summarised or are they "heterogeneous"? Network meta-analysis (also called multiple-treatments meta-analysis) additionally uses "indirect evidence". For example, if in schizophrenia there were trials that compared olanzapine with quetiapine and trials that compared olanzapine with aripiprazole, but no trials comparing quetiapine with aripiprazole *directly*, we can estimate quetiapine versus aripiprazole *indirectly* from the other two *direct* comparisons (see Fig. 1). There are several strengths and added values of this approach: (a) the indirect evidence can fill in the gaps in the evidence matrix, which allows to come up with hierarchies of which drug is probably the best, second best, third best and so on. This information is urgently needed by guidelines, but cannot really

✉ Stefan Leucht
  Stefan.Leucht@tum.de

  Anna Chaimani
  annachaimani@gmail.com

  Andrea S. Cipriani
  andrea.cipriani@psych.ox.ac.uk

  John M. Davis
  jdavis@psych.uic.edu

  Georgia Salanti
  georgia.salanti@ispm.unibe.ch

1 Department of Psychiatry and Psychotherapy, Technische Universität München Klinikum rechts der Isar, Ismaningerstr. 22, 81675 Munich, Germany

2 Department of Hygiene and Epidemiology, University of Ioannina School of Medicine, University Campus Ioannina, 45110 Ioannina, Greece

3 Department of Psychiatry, Warneford Hospital, University of Oxford, Oxford OX3 7JX, UK

4 Psychiatric Institute, University of Illinois at Chicago, 1601 W. Taylor St., Chicago, IL 60612, USA

5 Maryland Psychiatric Research Center, Baltimore, MD, USA

6 Departments of Health Promotion and Human Behavior and of Clinical Epidemiology, Kyoto University Graduate School of Medicine/School of Public Health, Kyoto, Japan

7 Institute of Social and Preventive Medicine (ISPM) and Berner Institut für Hausarztmedizin (BIHAM), University of Bern, Finkenhubelweg 11, 3012 Bern, Switzerland

Fig. 1 Principle of the use of indirect evidence in network meta-analysis
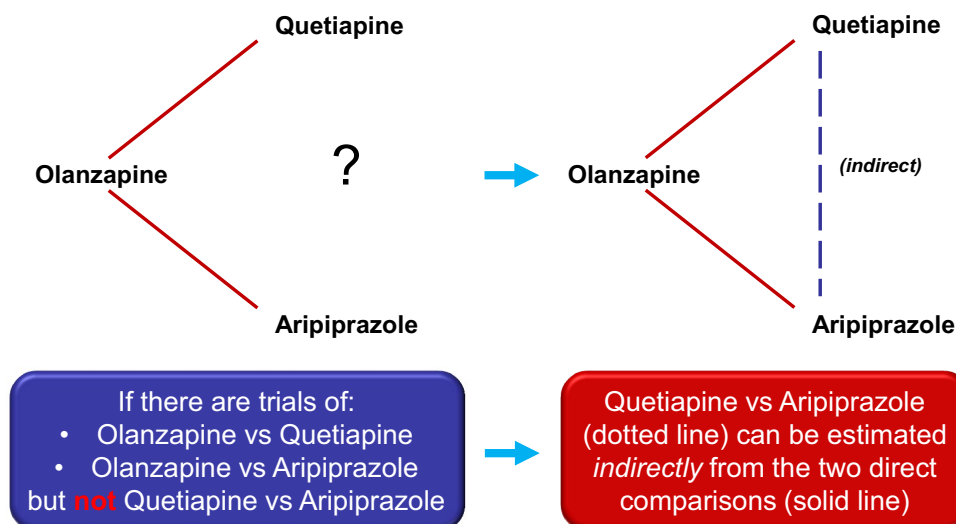
be provided by conventional meta-analysis (now sometimes also called "pairwise meta-analysis"). (b) Network meta-analysis can use all kinds of comparisons simultaneously—single antipsychotics versus placebo [11], head-to-head comparisons of new versus old antipsychotics [13], head-to-head comparisons of new drugs [15] —these separate types of comparisons could heretofore be only summarised in separate meta-analyses and viewed "impressionistically" together afterwards [14]. When the network is well connected and provides both direct (e.g. quetiapine vs. aripiprazole directly head-to-head) and indirect (e.g. quetiapine vs. aripiprazole via olanzapine) comparisons, they can be pooled together in the so-called mixed evidence, thus increasing statistical power and the precision of the estimates [3]. This use of the entire information also allows for more timely recommendations compared to conventional pairwise meta-analyses [16]. The underlying assumption of NMA is whether the indirect evidence validly estimates the differences between treatments. This issue is examined in several ways including statistical tests that compare the direct and indirect evidence for all comparisons where both are available [17].

Nevertheless, we feel that there are at least two major arguments why network meta-analysis and conventional pairwise meta-analyses should generally be considered the highest level of evidence (Fig. 2).

1. The first one is a simple, pragmatic argument: Nowadays, there are so many trials available, that it is simply impossible for a guideline team to read them all and to come up with an objective evaluation. For example, the latest network meta-analysis on antipsychotic drugs for schizophrenia comprised 212 blinded trials [12], and the last network meta-analysis on antidepressants for major depressive disorder 117 randomised-controlled
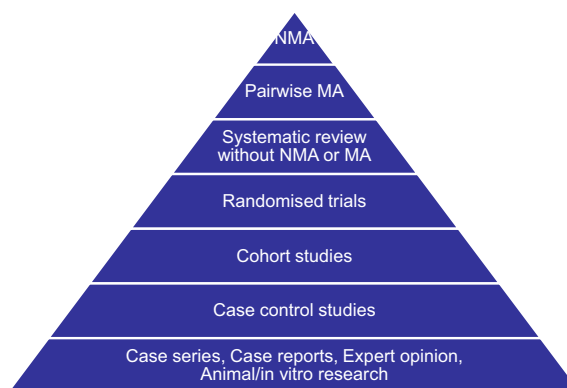
NMA = network (or multiple-treatments) meta-analysis, MA = meta-analysis, NMA and MA must have applied systematic review methods

Fig. 2 Proposed evidence hierarchy

trials [2]. Nobody can read all these articles and objectively "synthesise" them narratively. We have shown that abstracts from industry-sponsored trials are often biased, thus to read only the abstracts is not sufficient [9]. Actually, the avalanche of evidence is even a problem now for meta-analyses. In 2010, 11 meta-analyses were published per day, the same amount of randomised-controlled trials published three decades ago [1]. There are often several meta-analyses on the same or similar topics. But, their authors do not always come up with the same conclusions, and it is often unclear whether the reason are slightly different research questions, or different interpretation of the results [8]. We have therefore demanded to make a review of the existing systematic reviews mandatory [8].

2. The second argument is that science always has to start out from the ideal situation. Imagine 10 identical stud-

ies. There is no doubt that the pooled estimate of these 10 studies is better evidence than any of the single studies alone. The simple reason is that a bigger sample size increases the precision of the estimate, meaning that we can be more confident about the result. Consider this thought experiment: there is a trial with 10 participants of which seven responded to treatment and an identical trial with 1000 participants of which 700 responded. In both cases, the response rate is 70 %, but obviously, we would trust the large trial more. The same holds true for network meta-analysis. If all trials are identically designed, and the direct and the indirect evidence are "consistent," then there is no reason to not use it for complementing the evidence. Therefore, nothing generally speaks against considering network meta-analysis as the highest level of evidence. The problem is rather that often the world is not ideal. For example, it is well known from many medical fields that small trials tend to exaggerate treatment effects [4]. Therefore, the results of meta-analyses based on several trials can be completely reversed by one later published, large randomised-controlled trial [10, 19]. In mental health, for example, it has been shown that the results of meta-analyses only get stable once approximately, 1000 participants have been included in them [19]. In our experience, the results of network meta-analysis can also be distorted by small trials or by differences in other trial characteristics such as different study conditions, differences in patient inclusion criteria etc. But, these potential limitations (which may occur or not occur in a specific case) should not be used to a priori preclude network meta-analysis from the top of the evidence hierarchy. In medicine, we should always start from the theoretically best method, and all methods have limitations. For example, similar problems occur in randomised-controlled trials which are preferred by some guideline producers. Can we really assume that the patients included in them are similar enough that we can average the effects in both groups and compare them? The inclusion criteria usually leave a lot of room for variability leading to large standard deviations in psychiatric studies.

Therefore, in our opinion, systematic reviews based on network meta-analyses should generally be the highest level of evidence in treatment guidelines, but we need to assess them carefully and in certain situations (such as if a meta-analysis is mainly composed of small trials), later published well-designed, large randomised-controlled trials may indeed be preferred [6].

Nothing comes in complete black or white, but they come in shades of grey in the real world. It is therefore imperative for evidence users to critically appraise each piece of evidence, be it network meta-analysis, pairwise meta-analysis or randomised-controlled trial. One general problem is that publications on the level of evidence often omit the term "systematic review" before meta-analysis—probably only because otherwise the term gets very they long—but a systematic review process must always be implied because without it, any meta-analyses can be useless and should be disregarded. Checklists to assess the quality of systematic reviews such as the AMSTAR instrument exist, but they only check the methodological quality of a systematic review, for example, whether there was a systematic literature search or whether publication bias been investigated [18]. They do not examine the quality and content of the included studies, which should be assessed with the risk of bias tool (bearing in mind the risk of "garbage in garbage out"). It would be laudable if guideline authors could reassess the included studies themselves, but this requires a lot of expert knowledge, it is time consuming, and it opens the doors for selection bias. We would therefore favour the general application of the GRADE approach [7] which should be ideally already applied by the original systematic review authors, and for which extensions to network meta-analysis has been developed and should be endorsed [14].

Compliance with ethical standards

References

1. Bastian H, Glasziou P, Chalmers I (2010) Seventy-five trials and eleven systematic reviews a day: How will we ever keep up? PLoS medicine 7:e1000326
2. Cipriani A, Furukawa TA, Salanti G, Geddes JR, Higgins JP, Churchill R, Watanabe N, Nakagawa A, Omori IM, McGuire H, Tansella M, Barbui C (2009) Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet 373:746–758

3. Cipriani A, Higgins JP, Geddes JR, Salanti G (2013) Conceptual and technical challenges in network meta-analysis. Ann Intern Med 159:130–137

4. Dechartres A, Trinquart L, Boutron I, Ravaud P (2013) Influence of trial sample size on treatment effect estimates: meta-epidemiological study. BMJ 346:f2304

5. Evidence-Based Medicine Working G (1992) Evidence-based medicine. A new approach to teaching the practice of medicine. JAMA 268:2420–2425

6. Furukawa T (2004) Meta-analyses and megatrials: Neither is the infallible, universal standard. Evi Based Mental Health 7:34–35

7. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schunemann HJ, Group GW (2008) Grade: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 336:924–926

8. Helfer B, Prosser A, Samara MT, Geddes JR, Cipriani A, Davis JM, Mavridis D, Salanti G, Leucht S (2015) Recent meta-analyses neglect previous systematic reviews and meta-analyses about the same topic: a systematic examination. BMC Med 13:82

9. Heres S, Davis J, Maino K, Jetzinger E, Kissling W, Leucht S (2006) 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 163:185–194

10. Lelorier J, Grégoire G, Benhaddad A, Lapierre J, Derderian F (1997) Discrepancies between meta-analyses and subsequent large randomized, controlled trials. N Engl J Med 337:536–542

11. Leucht S, Arbter D, Engel RR, Kissling W, Davis JM (2009) How effective are second-generation antipsychotic drugs? A meta-analysis of placebo-controlled trials. Mol Psychiatry 14:429–447

12. Leucht S, Cipriani A, Spineli L, Mavridis D, Orey D, Richter F, Samara M, Barbui C, Engel RR, Geddes JR, Kissling W, Stapf MP, Lassig B, Salanti G, Davis JM (2013) Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. Lancet 382:951–962

13. Leucht S, Corves C, Arbter D, Engel RR, Li C, Davis JM (2009) Second-generation versus first-generation antipsychotic drugs for schizophrenia: a meta-analysis. Lancet 373:31–41

14. Leucht S, Kissling W, Davis JM (2009) Second-generation antipsychotic drugs for schizophrenia: Can we resolve the conflict? Psychol Med 39:1591–1602

15. Leucht S, Komossa K, Rummel-Kluge C, Corves C, Hunger H, Schmid F, Schwarz S, Davis JM (2009) A meta-analysis of head to head comparisons of second generation antipsychotics in the treatment of schizophrenia. Am J Psychiatry 166:152–163

16. Rouse B, Cipriani A, Shi Q, Coleman AL, Dickersin K, Li T (2016) Network meta-analysis for clinical practice guidelines: a case study on first-line medical therapies for primary open-angle glaucoma. Ann Intern Med 164:674–682

17. Salanti G, Del Giovane C, Chaimani A, Caldwell DM, Higgins JP (2014) Evaluating the quality of evidence from a network meta-analysis. PLoS ONE 9:e99682

18. Shea BJ, Hamel C, Wells GA, Bouter LM, Kristjansson E, Grimshaw J, Henry DA, Boers M (2009) Amstar is a reliable and valid measurement tool to assess the methodological quality of systematic reviews. J Clin Epidemiol 62:1013–1020

19. Trikalinos TA, Churchill R, Ferri M, Leucht S, Tuunainen A, Wahlbeck K, Ioannidis JPA (2004) Effect sizes in cumulative meta-analyses of mental health randomized trials evolved over time. JClinEpidemiol 57:1124–1130