REVIEW

Critical Issues on the Use of Network Meta-analysis in Psychiatry

Ayşegül Yildiz, MD, Eduard Vieta, MD, PhD, Christoph U. Correll, MD, Mateusz Nikodem, MSc, and Ross J. Baldessarini, MD

Abstract: The need to support clinical decision making and cost-effectiveness analyses in medicine, despite a dearth of head-to-head treatment comparisons, has encouraged the development of methods enabling indirect comparisons of treatment alternatives, including network meta-analysis (NMA). Valid application of NMA requires close similarity of compared trials, including their design, patient characteristics, and methods of diagnosis and symptomatic assessment. When biological or other objective measures of outcomes are not available, as is the case in psychiatric disorders, subtle differences in characteristics of trials or participants may lead to unrecognized incoherence within a network and thus to inconsistent results. By considering comparative-efficacy analyses of psychotropic drugs in major psychiatric disorders as working examples, we underscore the risks of violating the fundamental transitivity assumption in the context of NMA and suggest precautions for creating a coherent network. We conclude that with thoughtful and critical application, NMA can add useful information concerning the comparative benefits, risks, and costs of specific treatments in psychiatry.

Keywords: antimanic, drug treatment, meta-analysis, multiple treatments, network, psychotropic

Cost-effectiveness research is defined as "the synthesis of evidence that compares the benefits and harms of therapeutic alternatives for a clinical condition" and is used to guide decisions for selecting medical treatments.[1] In general, well-designed and well-implemented, head-to-head, randomized, controlled trials (RCTs) provide the most rigorous research evidence on the relative efficacy and risks of different interventions. Such direct, head-to-head comparisons of treatments, however, are uncommon in clinical medicine, including psychiatry.[2] This reality may, in part, reflect the commercial nature of most modern drug trials, in which a demonstrated inferiority to a competing product would be an undesirable outcome. The paucity of such head-to-head RCTs has encouraged development of a method that enables indirect comparisons among competing interventions by combining findings from separate trials involving a common comparator.[3–5] This idea of "borrowing strength" from indirect evidence was proposed by Higgins and Whitehead[6] in 1996 but has become popular only recently.[7] Terms applying to approaches using evidence synthesis include *network meta-analysis* (NMA), *mixed treatment comparisons meta-analysis*, and *multiple treatments meta-analysis*.[3–7] If these data-pooling methods are to be useful and broadly applicable, it is essential that clinical decision makers are familiar with what is required, with the associated advantages and disadvantages, and with the underlying principles and assumptions.[5] Accordingly, we present a brief overview of the topic.

From Department of Psychiatry, Dokuz Eylül University, Izmir, Turkey (Dr. Yildiz); International Consortium for Bipolar Disorder Research, McLean Hospital, Belmont, MA (Drs. Yildiz, Vieta, and Baldessarini); Bipolar Disorders Program, Institute of Clinical Neuroscience, Hospital Clinic, University of Barcelona, IDIBAPS, CIBERSAM, Barcelona, Catalonia, Spain (Dr. Vieta); Recognition and Prevention Program, Department of Psychiatry, Zucker Hillside Hospital, New York, NY (Dr. Correll); Faculty of Applied Mathematics, AGH University of Science and Technology, Krakow, Poland (Mr. Nikodem); Department of Psychiatry, Harvard Medical School, Boston, MA (Dr. Baldessarini).

Original manuscript received 2 March 2013; accepted for publication subject to revision 22 April 2013; revised manuscripts received 30 Aug 2013 and 4 Nov 2013.

Supported, in part, by educational and research grants from the Depression and Bipolar Disorder Foundation (Dr. Yildiz), Polish Ministry of Science and Higher Education (Mr, Nikodem), and Bruce J. Anderson Foundation and the McLean Private Donors Psychopharmacology Research Fund (Dr. Baldessarini).

Correspondence: Dr. A. Yildiz, Depression & Bipolar Disorders Foundation/ Duygudurum Vakfi-DUVAK, Yenikale Mahallesi, Ogretmen Zakir Karatas Sokak No. 3/3, Narlidere, Izmir, Turkey. Email: agul_yildiz@hotmail.com

DOI: 10.1097/HRP.0000000000000025

OVERVIEW OF NETWORK META-ANALYSIS

Assumptions

A fundamental assumption in all meta-analysis is that either the true treatment effect is constant across trials (fixed effects) or that the trial-specific treatment differences are from a common distribution (random effects).[4] NMA further assumes and requires that the true AB difference in direct A versus B trials be identical to, or at least from the same common distribution as, the AB difference estimated from pooled data (i.e., combined data from trials of treatment A versus C and B versus C).[4] It is as if all trials had examined all treatments of

interest but that, in each trial, results for all but two or three treatments had been randomly lost.[5] The NMA-specific assumption here is that the relative effects of treatments will be similar across the entire set of trials, irrespective of which treatments were actually evaluated, or that the trials are "exchangeable" in their patient samples, designs, conduct, and outcome measures. That is, one can estimate the AB difference, via the common comparator C, through A versus C trials and B versus C trials. This approach assumes that there is either no imbalance in the distribution of effect-modifiers across the different types of treatment comparisons or that no effect-modifiers exist.[8] Such comparability is called *transitivity* of effects between compared treatments[4,5,8]—and cannot be fully validated by the data themselves.[4,5,8] Since it relies heavily on the circumstances of each set of trials and their patient inclusion criteria, well-informed clinical judgment is essential.[4] Consistency and coherence in the context of NMA refer to the similarity of treatment effects estimated by direct and indirect evidence.[7] It is noteworthy that, whereas the similarity assumption relates only to indirect comparisons, the consistency assumption applies when there is both direct and indirect evidence for a given treatment comparison.[8]

Advantages and Promises

In contrast to traditional, pairwise meta-analysis, which offers a single effect estimate between two competing treatments, NMA allows for more data to be incorporated into analyses, for enhanced precision, and for both indirect and multiple comparisons, with effect estimates for all possible pairwise comparisons and a probability-based rank ordering among them.[1,8] A network can also promote future research by depicting existing direct and indirect evidence for comparisons of interest.[7,9] Moreover, NMA applications within a Bayesian framework can accommodate complex research questions by simultaneously considering several outcomes or by allowing the incorporation of expert opinion in the form of probability-based prior distributions.[3,8] Finally, in situations where indirect evidence can eradicate trial-specific biases that are sometimes not identifiable in head-to-head meta-analyses, NMA estimates might be preferable.[8]

Limitations and Risks

Heterogeneity is an artificial or real between-study discrepancy in a treatment effect across different trials, and it can be determined via statistical, clinical, or methodological means. *Statistical heterogeneity* usually occurs when estimates of a treatment effect (e.g., odds ratio or relative risk) that has been obtained from different studies vary more than would be expected by chance. *Clinical heterogeneity* ensues when there are significant differences between individual studies in terms of the patients sampled, the treatment protocols used, or the settings in which the trials were undertaken. Examples in psychiatry show that a number of these variables are subject to change over time because of evolving treatment alternatives and environmental contexts, and more widespread geographical distribution of trial sites.[10,11] Finally, *methodological heterogeneity* refers to the study design and analysis—for example, randomization versus no randomization, or monotherapy versus add-on or combination treatments.[3] Statistical heterogeneity exists between trials making the same comparison; in the context of NMA, however, evidence of inconsistency lies between different types of comparisons that form evidence cycles (within a network).[12]

Although randomization is maintained within trials, it does not hold across the entire set of included trials, since patients are not randomized to different trials. Consequently, systematic differences in study characteristics or patient characteristics may exist across trials. If these characteristics may potentially influence treatment effects—in other words, if they are effect-modifiers—then an imbalance in their distribution across trial sets presents a potential problem.[8] In the context of NMA, three types of variation in treatment effects can occur: (1) true within-study variation of treatment effects, which can be detected only with individual patient-level data or the reporting of subgroups, (2) true between-study variation in treatment effects for a specific treatment comparison, and (3) true between-comparison variation in treatment effects.[8] In contrast to a single, pooled effect estimate that is obtained in a standard pairwise meta-analysis, NMA yields more information. Thus, between-study discrepancies and inconsistencies between pairwise comparisons in the context of NMA may affect several pooled effect estimates.[13] In situations where certain effect-modifiers distribute unevenly across sets of comparisons, the transitivity assumption no longer holds. Accordingly, employment of indirect comparisons is no longer possible. In such situations, where discrepancies between sets of comparisons exist, statistical checking of consistency also fails to identify these discrepancies, typically because of insufficient power. Consequently, expert clinical and epidemiological judgment must be employed very early on to assure similarity of the trials to be included. In addition, pre and post hoc statistical evaluations of potential effect-modifiers remain invaluable for a coherent network analysis.[4,5,9]

CHALLENGES OF NMA APPLICATIONS IN PSYCHIATRIC DISORDERS

The availability of many treatment options without the proven superiority of one drug over another makes NMA appealing to psychiatry. We searched the PubMed database to identify NMAs of psychotropic agents by using "multiple treatments meta-analysis, network meta-analysis, or mixed treatment comparisons meta-analysis" and "antipsychotic(s), antimanic(s), antidepressant(s), psychotropic, bipolar, mania, depression, schizophrenia, panic disorder, and addiction" as search terms through June 2013. We identified nine NMAs on the comparative efficacy of psychotropic drugs in major psychiatric disorders. Gartlehner and colleagues[13] were the first, in 2008, to use indirect comparisons to assess the benefits and risks of second-generation antidepressants

(SGAs), and they later updated their findings.[14] Subsequently, Cipriani and colleagues reported on the comparative efficacy and acceptability of modern antidepressants in major depression,[15] followed by a similar study of antimanic drugs in bipolar mania.[16] In 2012, Mills and colleagues[17] employed NMA to assess comparative effectiveness of nicotine replacement therapy, varenicline, and bupropion for smoking cessation. The US Department of Health and Human Services initiated an extensive NMA on the comparative effectiveness of first-generation versus second-generation antipsychotics in schizophrenia, which was published in 2012 by Hartling and colleagues.[18] An investigator-initiated NMA appeared in 2013 on the comparative efficacy and safety of the 2 most common first-generation antipsychotics and the 13 most common SGAs for treating schizophrenia.[19] This methodologically sound NMA, besides providing measures on efficacy and all-cause discontinuation rates, also analyzed electrocardiographic measures and quantifiable data on sedation, weight gain, extrapyramidal effects, and prolactin concentrations.[19] A comparative-efficacy assessment of antimanic treatments based on NMA also was recently updated.[20] Rather than reviewing the results of the combined use of direct and indirect evidence in each of these reports, we consider them here as working examples for discussing the most prominent pitfalls and challenges associated with NMA in psychiatry. Special emphasis is given to antimanic treatment comparisons.

Given both the substantial responses to placebo in most psychiatric disorders and the requirements of drug regulatory authorities to compare new treatments to placebo, RCTs in major psychiatric disorders routinely involve placebo controls but rarely include active treatment arms that would allow direct, head-to-head comparisons.[14,16,18,19] Further, even if an active comparator is used, the sample size is calculated to test the experimental treatment's superiority to placebo—and not the relative superiority or noninferiority of the two active treatments to one another. Thus, the comparator serves merely to establish the trial's "assay sensitivity." Data networks for psychotropic interventions consequently configure like radiating stars or polygons, for which quantifying the amount of incoherence between indirect and direct evidence is challenging.[1,7,9]

The first NMA on the comparative efficacy and acceptability of antimanic treatments considered all RCTs against placebo or other treatments as monotherapy but also included trials involving add-on and combination designs.[16] In the context of NMA, this approach assumes, in effect, that the relative effects of all the treatments in the studies with single-agent and add-on or combination designs are homogeneous and from the same common distribution.[5] However, the comparability of baseline characteristics in terms of disease severity and patients' potential for being drug or placebo responders is debatable in add-on versus single-agent trials of bipolar mania.[3,6] Add-on trials in acute mania, for example, typically involve patients who failed to respond adequately to lithium or valproate. Consequently, in a scenario where lithium response operates as an effect-modifier, an imbalance in the distribution of lithium responders versus nonresponders across trial sets would be inevitable.[16] Indeed, not only may the effects of placebo treatment differ substantially, but so may the effect of antimanic treatment in add-on versus single-agent trials. As an illustrative example, two add-on trials of risperidone to ongoing mood stabilizers indicated a faster[21] and greater decline[22] in the Young Mania Rating Scale scores with add-on risperidone relative to add-on placebo.[16] This initial network not only included, but also combined, findings from add-on and combination treatments with data from monotherapy trials (e.g., the haloperidol node included data derived from trials not only of haloperidol as monotherapy but also of haloperidol plus lithium or divalproex).[16] Additionally, chemically similar drugs were considered conjointly in the same node (e.g., data from RCTs involving paliperidone as monotherapy or add-on were considered together with data for risperidone, again as monotherapy or add-on), apparently based on chemical similarities that may, or may not, translate to pharmacological comparability.[9]

Consequently, whereas the network described above reported haloperidol superior to seven other agents (aripiprazole, asenapine, carbamazepine, lithium, quetiapine, valproate, and ziprasidone) and both olanzapine and risperidone superior to two other treatments (valproate and ziprasidone),[16] an updated NMA that considered only single-agent antimanic treatments failed to confirm most of those results.[20] In this recent NMA, which involved 14,256 manic patients randomized to one of 18 active treatments or placebo, point estimates based on standardized mean differences ranged from 0.32 to 0.66 for the 12 effective, widely prescribed antimanic drugs. These results indicated that no treatment was superior to any other except for risperidone versus aripiprazole and valproate. A sensitivity analysis by drug class also yielded comparable effect estimates for haloperidol, SGAs, and mood stabilizers (carbamazepine, lithium, or valproate).[20] These two NMAs of antimanic treatments illustrate how differences in network construction and evaluation may result in different rankings by efficacy based on point estimates and the surface under the cumulative ranking curve (SUCRA), and also in different effect estimates for the competing interventions based on standardized mean differences.[16,20]

In the context of NMA, variation in responses to placebo within or between antimanic-treatment trial sets of a particular design calls for further attention. Add-on trials of acute mania typically involve patients who have been unresponsive to previous treatments; hence, they have special patient characteristics and involve different placebo-associated responses. Therefore, combining add-on trials with monotherapy trials in an antimanic treatment network where the "common comparator" for most indirect comparisons is placebo may jeopardize the transitivity assumption.[7,23] Besides, in the case of acute mania, the bulk of accumulated evidence is from single-agent RCTs—unlike bipolar depression or maintenance trials, where the research question is frequently

whether combination treatments have effects superior to monotherapy. Accordingly, a network based solely on monotherapy trials seems to be more appropriate for antimanic treatment comparisons, whereas RCT sets of single-agent, add-on, or combination treatments may be more suitable for bipolar depression or bipolar maintenance trials. Such decisions regarding NMA applications in psychiatry should be based on expert clinical judgment, the research question of interest, and the data structure of the available evidence, with a special emphasis on the similarity of patient and trial characteristics.

Another debate on the use of NMA in psychiatry concerns the combination of placebo-controlled and head-to-head trial evidence in the same network. For example, the first NMA of antidepressant drugs categorized all relevant placebo-controlled and head-to-head RCTs as monotherapy, and reported no clinically important differences between antidepressants that were being used for treating major depression.[13] Another NMA of the same antidepressant drugs, however, included only head-to-head RCTs and reported mirtazapine and venlafaxine as the most efficacious options for treating depression, and duloxetine, fluvoxamine, paroxetine, and reboxetine as the least efficacious.[15] Given that most available evidence on the efficacy of antidepressants is as measured against placebo treatment, the exclusion of placebo-controlled RCTs from the latter NMA might have altered the results.[13,15] With substantial variation in responses to placebo, on the one hand, and with the bulk of the available evidence involving placebo controls, on the other, the most suitable approach for evaluating psychotropic drugs via NMA remains a matter for debate. Nonetheless, the regulatory requirements for placebo controls—which has the consequence that the data on the treatment of major psychiatric disorders regularly include comparisons to placebos—suggest that these comparisons capture the primary question of clinical interest. Accordingly, most NMAs of psychotropic drugs involve, for better or worse, a placebo node.[16,19,20] Significantly, in the most recent NMA of antimanic treatments, a sensitivity analysis involving the exclusion of the placebo node artificially affected and skewed the results[20]—just as in the case of antidepressant networks. The same NMA also demonstrated that most critical, potential effect-modifiers (presence of psychotic features, mixed-state diagnosis, all-cause discontinuation rates, and publication year, all operating across the placebo-controlled RCTs of antimanic treatments)[23] do not survive when head-to-head trials and indirect comparisons are embraced in the context of NMA.[20] The analysis thus illustrates how indirect evidence within the framework of an NMA is potentially capable of eradicating some of the subtle, trial-specific biases that operate within the milieu of placebo-controlled RCTs.[8,20]

A further assumption of NMA concerns the scale of measurement. Indirect- and multiple-comparison models assume that treatment effects add together, with the consequence that the relative effect of treatments A versus C can be predicted through the effects of A versus B and B versus C.[5] This model assumes the choice of an appropriate measure of effect (e.g., log odds ratio, relative risk, or risk difference). The point is a critical one for NMAs in psychiatry, where outcome is typically assessed by rating scales rather than objective biological measures. For example, we can question whether a treatment effect in an A versus B trial that is measured by the Brief Psychiatric Rating Scale (range: 18–126) is exchangeable with one in a B versus C trial measured by the Young Mania Rating Scale (range: 0–60), and whether such a comparison will yield a valid estimate of the the A versus C difference—even if the analysis is adjusted by standard deviation or variance.[16,20] In the given example, the former scale is designed to assess psychotic and agitated behavior,[24] whereas the latter is designed to assess manic symptoms.[25] Although the use of dimensionless effect-size measures such as standardized mean difference technically enables such comparisons, it may be prudent—given the inherent subjectivity of assessments via rating scales in NMAs of psychotropic drugs—to consider RCTs whose assessment methods similarly target the actual psychopathology.

One further consideration is the range of effect-size estimates. For most psychotropic agents in clinical use,[16,19,20,26] the range is typically 0.32 to 0.66. This limited range of possible effect sizes requires that analyses of comparative efficacy be sensitive enough to capture small differences among treatments. At the other extreme, such a narrow range of observed point estimates may generate false-positive findings—particularly in the presence of unrecognized sources of bias or unevenly distributed effect-modifiers across trial sets. As noted above, undetected bias can lead to undetected incoherence—and to misleading results.[1,27]

Finally, as with the standard pairwise meta-analysis, NMA may be affected by publication and sponsorship bias.[1] A critical NMA-based report on antidepressants considered data from 74 FDA-registered, placebo-controlled trials of 12 antidepressants and their 51 matching publications to estimate the effect sizes for 66 possible pairwise comparisons of these drugs.[1] Independent meta-analyses of the FDA-registered data found smaller treatment effects for all drugs and a decrease in effect sizes, ranging from 0.10 or 0.11 for fluoxetine and controlled-release paroxetine to 0.39 or 0.41 for mirtazapine and nefazodone.[1] The estimates differed in absolute values by at least 100% for 30 of 66 pairwise comparisons, and by at least 50% for 44 of the same.[1] In an incomplete network that excluded only 4 unpublished trials of mirtazapine, the probability of mirtazapine ranking first was 80.6% compared to 7.3% in the complete network that included all 74 FDA-registered trials (including the 4 unpublished trials). This outcome highlights the potentially detrimental effect of publication bias.[1]

With regard to sponsorship bias, a review of head-to-head, SGA studies found that the reported overall outcome was in favor of the sponsor's drug in 90% of cases.[28] Studies on the same drugs but with different trial sponsors typically

resulted in contradictory conclusions. For example, for the treatment of acute mania, a study that was sponsored by the manufacturer of olanzapine found a slightly better response with olanzapine than with divalproex in 3 weeks,[29] whereas another trial sponsored by the manufacturer of divalproex found similar antimanic effects with both treatments in 12 weeks.[30] Dosing strategies, patient populations, or trial durations might have affected the results.[28] Analysts conducting NMAs on competing psychotropic interventions should be aware of the possibility of subtle sponsorship bias.[9] Recent NMAs—one on antipsychotic treatments in schizophrenia and another on antimanic treatments for bipolar acute mania—provide examples of dealing with subtle biases (which potentially result from dosing strategies, allocation concealment, or analytic approaches) by applying sensitivity and meta-regression analyses.[19,20]

FUTURE INSIGHTS

The careful and critical application of NMAs holds considerable promise to guide decisions on clinical care and mental health policies. NMA is as valid as pairwise meta-analysis, provided that the quantity and quality of RCT data enables the construction of a coherent network.[8] We believe that recent NMAs on antipsychotic drugs in schizophrenia and antimanic treatments for bipolar mania show how NMA's technical advantages can be used in psychiatry.[19,20]

Taking into account the above-mentioned considerations, we propose the following: (1) In constructing coherent networks for the study of major psychiatric disorders, prior and up-front expert clinical and epidemiological judgment is required to ensure the inclusion of similar trials as determined by the research question of interest and the target population. (2) NMAs in psychiatry should explicitly and pragmatically consider and evaluate a broad range of factors (e.g., patient, illness, treatment, design, trial conduct, and setting) as potential effect-modifiers. (3) Pharmaceutical regulatory agencies should potentially consider the requirement of active controls in addition to placebo controls.[31] If NMAs in psychiatry become standardized and established, at some point it might even be possible to forgo placebo-controlled trials in designated therapeutic areas. For instance, once compelling evidence on the effect of a drug against placebo has accumulated, comparative trials using a second agent can provide an indirect estimate of a test drug's effect against placebo.[32] (4) Bayesian-oriented NMA applications allow the inclusion of various factors within the same decision-analytic, cost-effectiveness model.[3,4,7,9] The application of such models for synthesizing qualitative research in psychiatry may provide further information about complex interventions consisting of several interacting components and also about questions concerning translational and implementation science—provided that standardized data on such measures of interest are available.[33] These NMA applications are likely to play a major role in future health policy decisions. For example, NMAs may improve our knowledge about a treatment's effect on quality of life, total life-years gained, and cognition; on its neuro-protective versus neurodetrimental effects; on its role in preventing suicide; and on the associated risk of adverse neurological and general medical effects.[3,4,7,9]

Declarationof interest: Dr. Yildiz has received research grants from or served as a consultant to, or speaker for, Abdi Ibrahim, Actavis, Ali Raif, AstraZeneca, Bristol-Myers Squibb, Janssen-Cilag, Lundbeck, Pfizer, Sanofi-Aventis, and Servier. Dr. Correll has received research grants from, been a consultant or adviser to, or received honoraria from Actelion, Alexza, AstraZeneca, Bristol-Myers Squibb, Cephalon, Eli Lilly, Gerson Lehrman Group, Genentech, IntraCellular Therapies, Janssen/Johnson & Johnson, Lundbeck, Medavante, Medscape, Merck, Novartis, Ortho-McNeill/Janssen/Johnson & Johnson, Otsuka, Pfizer, ProPhase, Roche, Sunovion, Takeda, Teva, and Vanda. Professor Vieta has received research grants from or served as a consultant, adviser, or speaker for companies that include AstraZeneca, Bristol-Myers Squibb, Eli Lilly, Forest Research Institute, Gedeon Richter, GlaxoSmithKline, Janssen-Cilag, Jazz, Johnson & Johnson, Lundbeck, Merck, Novartis, Otsuka, Pfizer, Roche, Qualigen, Sanofi-Aventis, Servier, Schering-Plough, Takeda, and Teva.

REFERENCES

1. Trinquart L, Abbé A, Ravaud P. Impact of reporting bias in network meta-analysis of antidepressant placebo-controlled trials. PLoS One 2012;7(4):e35219.
2. Song F, Loke YK, Walsh T, Glenny AM, 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.
3. Ades AE, Sculpher M, Sutton A, et al. Bayesian methods for evidence synthesis in cost-effectiveness analysis. Pharmacoeconomics 2006;24:1–19.
4. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004;23:3105–24.
5. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005;331:897–900.
6. Higgins JP, Whitehead A. Borrowing strength from external trials in a meta-analysis. Stat Med 1996;15:2733–49.
7. Salanti G, Marinho V, Higgins JPT. Case study of multiple-treatments meta-analysis demonstrates that covariates should be considered. J Clin Epidemiol 2009;62:857–64.
8. 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.
9. Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomized trials. Stat Methods Med Res 2008;17:279–301.
10. Correll CU, De Hert M. Antipsychotics for acute schizophrenia: making choices. Lancet 2013;382:919–20.
11. Vieta E, Pappadopulos E, Mandel FS, Lombardo I. Impact of geographical and cultural factors on clinical trials in acute mania: lessons from a ziprasidone and haloperidol placebo-controlled study. Int J Neuropsychopharmacol 2011;14:1017–27.

12. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. J Am Statist Assoc 2006;101:447–59.
13. Gartlehner G, Gaynes BN, Hansen RA, et al. Comparative benefits and harms of second-generation antidepressants: background paper for the American College of Physicians. Ann Intern Med 2008;149:734–50.
14. Gartlehner G, Hansen RA, Morgan LC, et al. Comparative benefits and harms of second-generation antidepressants for treating major depressive disorder: an updated meta-analysis. Ann Intern Med 2011;155:772–85.
15. Cipriani A, Furukawa TA, Salanti G, et al. Comparative efficacy and acceptability of 12 new-generation antidepressants: a multiple-treatments meta-analysis. Lancet 2009;373:746–58.
16. Cipriani A, Barbui C, Salanti G, et al. Comparative efficacy and acceptability of antimanic drugs in acute mania: multiple-treatments meta-analysis. Lancet 2011;378:1306–15.
17. Mills EJ, Wu P, Lockhart I, Thorlund K, Puhan M, Ebbert JO. Comparisons of high-dose and combination nicotine replacement therapy, varenicline, and bupropion for smoking cessation: a systematic review and multiple treatment meta-analysis. Ann Med 2012;44:588–97.
18. Hartling L, Abou-Setta AM, Dursun S, Mousavi SS, Pasichnyk D, Newton AS. Antipsychotics in adults with schizophrenia: comparative effectiveness of first-generation versus second-generation medications: a systematic review and meta-analysis. Ann Intern Med 2012;157:498–511.
19. Leucht S, Cipriani A, Spineli L, et al. Comparative efficacy and tolerability of 15 antipsychotic drugs in schizophrenia: a multiple-treatments meta-analysis. Lancet 2013;382:951–62.
20. Yildiz A, Nikodem M, Vieta E, Correll CU, Baldessarini RJ. A network meta-analysis on comparative efficacy and all-cause discontinuation of antimanic treatments in acute bipolar mania. Psychol Med 2014:18 July [Epub ahead of print].
21. Yatham LN, Grossman F, Augustyns I, Vieta E, Ravindran A. Mood stabilisers plus risperidone or placebo in the treatment of acute mania. International, double-blind, randomised controlled trial. Br J Psychiatry 2003;182:141–7.
22. Sachs GS, Grossman F, Ghaemi SN, Okamoto A, Bowden CL. Combination of a mood stabilizer with risperidone or haloperidol for treatment of acute mania: a double-blind, placebo-controlled comparison of efficacy and safety. Am J Psychiatry 2002;159:1146–54.
23. Yildiz A, Vieta E, Tohen M, Baldessarini RJ. Factors modifying drug and placebo responses in randomized trials for acute mania. Int J Neuropsychopharmacol 2011;14:863–75.
24. Overall JE, Gorham DR. Brief Psychiatric Rating Scale (BPRS): recent developments in ascertainment and scaling. Psychopharmacol Bull 1988;24:97–9.
25. Young RC, Biggs JT, Ziegler VE, Meyer DA. Rating scale for mania: reliability, validity and sensitivity. Br J Psychiatry 1978;133:429–35.
26. Yildiz A, Vieta E, Leucht S, Baldessarini RJ. Efficacy of antimanic treatments: meta-analysis of randomized controlled trials. Neuropsychopharmacology 2011;36:375–89.
27. 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.
28. Heres S, Davis J, Maino K, Jetzinger E, Kissling W, Leucht S. 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.
29. Tohen M, Baker RW, Altshuler LL, et al. Olanzapine vs. divalproex in the treatment of acute mania. Am J Psychiatry 2002;159:1011–7.
30. Zajecka JM, Weisler R, Sachs G, Swann AC, Wozniak P, Sommerville KW. Comparison of the efficacy, safety, and tolerability of divalproex sodium and olanzapine in the treatment of bipolar disorder. J Clin Psychiatry 2002;63:1148–55.
31. Vieta E, Cruz N. Head to head comparisons as an alternative to placebo-controlled trials. Eur Neuropsychopharmacol 2012;22:800–3.
32. Psaty BM, Lumley T, Furberg CD, et al. Health outcomes associated with various antihypertensive therapies used as first-line agents. A network meta-analysis. JAMA 2003;289:2534–44.
33. Kriston L, Yildiz A. Traditional and novel research synthesis approaches to support evidence-based treatment decisions. In: Yildiz A, Ruiz P, Nemeroff P, eds. The bipolar book: history, neurobiology, and treatment. Oxford; New York: Oxford University Press (forthcoming).