International Journal of Technology Assessment in Health Care, **24:2** (2008), 170–177.
Copyright © 2008 Cambridge University Press. Printed in the U.S.A.
DOI: 10.1017/S0266462308080240

Direct versus indirect comparisons: A summary of the evidence

Gerald Gartlehner
Danube University Krems

Charity G. Moore
University of Pittsburgh

Objectives: The aim of this study was to provide a comprehensive summary and interpretation of the current evidence on the use and validity of statistical methods to conduct indirect comparisons of treatment effects.
Methods: A narrative review was conducted.
Results: Well-conducted methodological studies provide good evidence that adjusted indirect comparisons can lead to results similar to those from direct comparisons. The internal validity of several statistical methods to conduct indirect comparisons, therefore, has been established. Meta-regression, logistic regression, or adjusted indirect comparisons should be the methods of first choice. Unadjusted indirect comparisons are always unacceptable. Deciding whether to combine direct and indirect evidence will be a matter of informed judgment based on the similarities and dissimilarities of populations and interventions. Unverifiable assumptions with respect to the similarity of compared studies and low power are serious limitations of indirect comparisons.
Conclusions: In the absence of sufficient head-to-head evidence, adjusted indirect comparisons can be viewed as additional analytical tools to determine the comparative efficacy and effectiveness of competing interventions. Researchers who use indirect comparisons need to keep the limitations in mind.

Keywords: Indirect comparisons, Adjusted indirect comparisons, Summary of the evidence

In recent years, health technology assessments (HTAs) have become a primary source of evidence for healthcare decision makers confronted with questions about the comparative effectiveness and safety of interventions. The usefulness of such reports, however, depends largely on the availability of primary research addressing a particular question of interest. To determine the best treatment option, decision makers and patients must be able to compare the net benefits of any competing interventions.

Head-to-head randomized controlled trials (RCTs) are considered the gold standard for evaluating the comparative efficacy, effectiveness, and safety of healthcare interventions. Many competing interventions, however, have never been compared directly in RCTs (18). Therefore, researchers who conduct HTAs often have to decide whether or not to use statistical methods to compare competing interventions indirectly.

Traditionally, researchers have dealt with the lack of direct evidence by qualitatively comparing effect sizes of competing interventions. For example, researchers often examine whether confidence intervals of different treatment effects overlap. In a drug class review on proton pump inhibitors (PPIs), researchers pooled esophagitis healing rates for each drug at 4 and 8 weeks, using PPI arms of studies with different comparisons (25). Authors presented pooled point estimates with confidence intervals for each drug in a table

We thank our colleagues Kathleen N. Lohr, PhD, and Sally Morton, PhD, from RTI-International for insightful comments on earlier drafts of this manuscript.

https://doi.org/10.1017/S0266462308080240 Published online by Cambridge University Press

and concluded, "Healing rates were similar and confidence intervals overlapped, indicating no significant differences between PPIs." Such qualitative assessments, although very common, are very susceptible to biases and confounders. A recent review of the Database of Abstracts of Reviews of Effects (DARE) found that 9.5 percent of meta-analyses used indirect comparisons (13). Few studies, however, carried out a formal statistical analysis to conduct adjusted indirect comparisons.

Increasingly, methods researchers have been evaluating the validity and reliability of various statistical methods to conduct indirect comparisons (2;5;6;13;23;39). A recent HTA, conducted by Glenny and colleagues (13) for the NHS R&D (National Health System Review and Dissemination) Health Technology Program, is the largest, most thorough attempt so far.

The objective of this study is to summarize the current state of the evidence on the pros and cons of indirect comparisons using study level data. We conducted a Medline search using "indirect comparisons," "adjusted indirect comparisons," and "network meta-analysis" as text words because no respective MeSH (Medical Subject Heading) terms were available. In addition, we hand-searched reference lists of relevant methods articles. We set forth to illustrate alternative approaches of indirect comparisons and highlight advantages and disadvantages of different methods to provide a summary for researchers who are considering indirect comparisons and readers who are critically appraising results of such methods. Table 1 summarizes the existing evidence, the following paragraphs discuss findings in more detail.

Table 1. Summary of the evidence

Situation	Evidence
What alternative approaches are available for indirect comparisons?	Validated approaches are meta-regression, fixed and random effects adjusted indirect comparisons, and logistic regression. More complex approaches such as network meta-analysis or Bayesian approaches have not been thoroughly evaluated.
Is there a preferred approach?	In simulation studies, meta-regression, adjusted indirect comparisons, and logistic regression provided unbiased results. Choice of fixed versus random effects models is scenario-dependent. More complex approaches have not been validated yet.
Are some approaches unacceptable?	In simulation studies, unadjusted indirect comparisons led to overly precise, biased results.
How many studies are needed for indirect comparisons?	Indirect comparisons can be conducted with as few as two studies. Point estimates in simulation studies were correct but the precision of estimates was low. Four times as many equally sized studies are necessary for indirect comparisons to have the same precision as direct comparisons.
How can results from direct and indirect comparisons be combined?	No evidence exists.
What if results from direct and indirect comparisons conflict?	No evidence exists.

WHAT ARE INDIRECT COMPARISONS?

A typical example of a body of evidence with insufficient information from head-to-head trials is depicted in Figure 1. The comparison of interest is the efficacy of two second-generation antidepressants for the treatment of major depressive disorder: sertraline versus venlafaxine. Only two small studies have compared these drugs directly (26;38) and rendered conflicting results. One study reported statistically significantly higher response rates (defined as an at least 50 percent improvement of symptoms on the Hamilton Depression Rating Scale) for patients treated with venlafaxine than with sertraline (83 percent versus 68 percent; $p = .05$) (26), the other trial found no differences in response between treatments (71 percent versus 71 percent; $p = .95$) (38). Both studies lasted 8 weeks and had no major methodological flaws. Based on the available evidence, therefore, it is impossible to assess reliably the comparative efficacy of these two agents.

In addition to the two direct head-to-head studies, however, seven placebo-controlled trials on sertraline (17;20;24;30–32;35) and four on venlafaxine (8;19;27;41) provide an additional body of evidence that can be used. Likewise, eighteen trials used fluoxetine as a common control intervention (1;3;4;7;9–12;21;28;29;33;34;36;37;42–44). Because of the inconclusive and sparse head-to-head data, indirect comparisons of these additional data around a common comparator can be taken into consideration as an additional analytic tool to achieve an estimate of the difference in efficacy between sertraline and venlafaxine.

METHODS OF INDIRECT COMPARISONS

Articles in peer-reviewed literature have used two general approaches to conduct such an indirect comparison: (i) unadjusted indirect comparison and (ii) adjusted indirect comparison. Figures 2 and 3 depict the general principles of these two methods.

https://doi.org/10.1017/S0266462308080240 Published online by Cambridge University Press

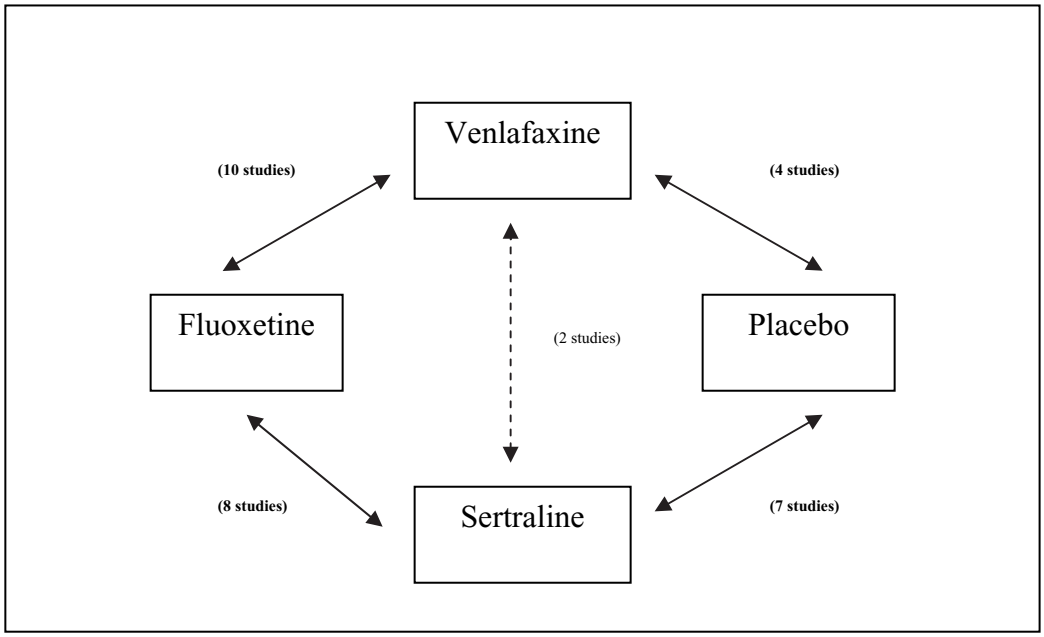


Figure 1. Body of evidence for the comparison of sertraline versus venlafaxine.

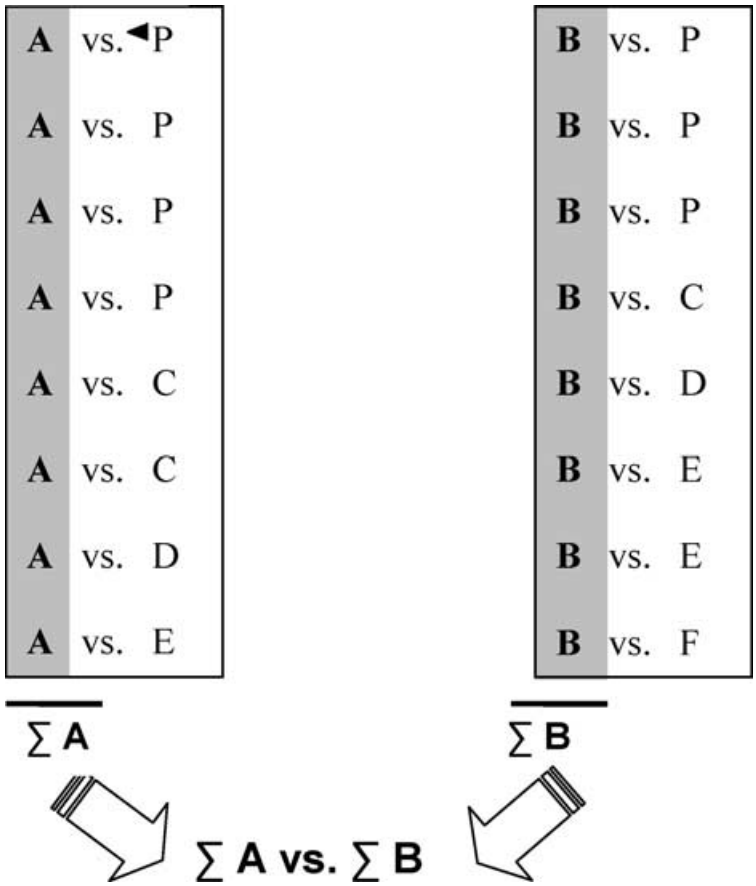


Figure 2. Unadjusted indirect comparisons.

https://doi.org/10.1017/S0266462308080240 Published online by Cambridge University Press

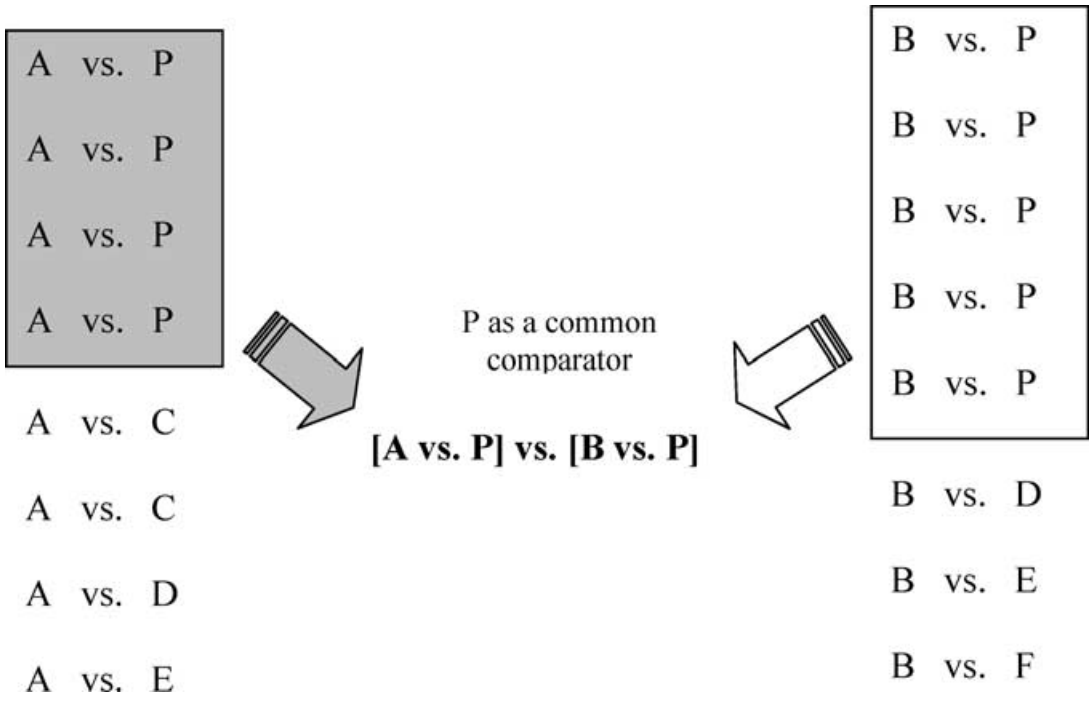


Figure 3. Adjusted indirect comparisons.

Unadjusted Indirect Comparisons

Unadjusted indirect comparisons make comparisons across trial arms, treating data as if they had come from one large trial. As illustrated in Figure 2, in a first step an unadjusted indirect comparison would combine the summary statistics for study arms involving intervention A to create a weighted overall summary measure. The same would be done for intervention B. In a second step, analysts would then compare these two overall summary measures. Data from the comparative arms are not used.

The principal methodological problem with this approach is that it negates the randomized nature of individual trials. Combining data across arms creates two large cohorts that may or may not be comparable in prognostic (or other potentially confounding) factors. Glenny et al. (13) termed this approach "the naïve method." When compared with direct estimates, this approach showed a high rate of discrepancies in unpredictable directions. In their simulation studies, this approach was unbiased but showed overly precise estimates as compared to other valid methods (adjusted indirect, logistic regression, and meta-regression). In addition, with such an approach, critics can legitimately question the exclusion of nonrandomized or observational studies. Therefore, unadjusted indirect comparisons should always be avoided.

Adjusted Indirect Comparisons

In adjusted indirect comparisons, analysts adjust the comparison of the interventions by the results of their direct comparison with a common control group (e.g., placebo or an active comparator), at least partially using the strength of the original RCTs (40). As illustrated in Figure 3, in a first step the effects of interventions A and B are compared with a common comparator (in our example P). The relative effects of A versus P are then compared with the relative effects of B versus P. For example, we can obtain an indirect estimate of the difference of benefits of sertraline versus venlafaxine by comparing the effect sizes of trials of sertraline versus placebo with those of trials of venlafaxine versus placebo. The underlying rationale is that, if the treatment effect of sertraline versus placebo is similar with that of venlafaxine versus placebo, then the comparative benefits of sertraline and venlafaxine will be similar as well. Of course, any common comparator can be used in place of placebo.

In adjusted indirect comparisons, data are *not* treated as if they had come from one large trial; rather, individual studies are maintained as the units of the analysis. As in a meta-analysis, investigators construct effect estimates of individual trials, weight them appropriately, and then combine them into a summary effect estimate. In adjusted indirect comparisons, then, they compare the summary effect estimates of the two sets of trials.

Conceptually, this follows a simple principle:

$$\theta_{AB} = \theta_{AP} - \theta_{BP}$$

where θ is the estimate of effect (e.g., log odds ratio, log relative risk, or means).

https://doi.org/10.1017/S0266462308080240 Published online by Cambridge University Press

Using our second-generation antidepressants example from above, the equation would read as follows:

$$\text{sertraline versus venlafaxine} = [\text{sertraline versus placebo}] - [\text{venlafaxine versus placebo}]$$

Important for the validity of the result of an adjusted indirect comparison is its dependence on the consistency of the treatment effect across the two different sets of trials. Theoretically, treatments of the two sets of trials should be interchangeable without leading to a different conclusion. Differences in any factors that can modify the treatment effect between the two sets of trials will lead to biased results. For example, patients treated with sertraline must have similar prognostic factors, such as severity of disease, prior depressive episodes, and others, as patients treated with venlafaxine. To some extent, this requires that study protocols of trials included for indirect comparisons have to be similar.

In practice, however, trials can vary in numerous ways: population characteristics, precise interventions, cointerventions, length of follow-up, loss to follow-up, and study quality. Some of these factors can easily be assessed and controlled for in regression models; others require assumptions that may not be verifiable. It is impossible to clearly determine to what extent these differences might bias results of indirect comparisons. Because indirect comparisons essentially constitute an observational study, residual confounding can always be present. The Cochrane Collaboration handbook for authors cautions that indirect comparisons may suffer all biases that can occur in observational studies (16).

DIFFERENT APPROACHES FOR THE ANALYSIS OF ADJUSTED INDIRECT COMPARISONS

In their methods study, Glenny et al. (13) evaluated four statistical approaches: (i) logistic regression, (ii) random effects meta-regression, (iii) fixed effects adjusted indirect comparisons [Bucher], and (iv) modified (random effects) adjusted indirect comparisons [Bucher] using simulations. (Note: we will refer to the method proposed by Bucher et al. as *adjusted indirect comparisons [Bucher]* (5) to distinguish this statistical method from the term "adjusted indirect comparisons" that is used as an umbrella term, as outlined earlier and depicted in Figure 3. Results of these simulation models indicated that, under ideal circumstances (i.e., no differences in prognostic factors existed among included studies), all four methods resulted in unbiased estimates of direct effects (13). For the simulated scenario, fixed effects methods (adjusted indirect comparisons [Bucher] and logistic regression) underestimated the true variability in the measure of association. In practice, the correctness of their standard errors would be impossible to test.

Although all four statistical approaches led to valid results, they differ in several characteristics that might guide the choice of a method. Logistic regression is limited to dichotomous outcomes and odds ratios as the measure of association. By contrast, meta-regression and adjusted indirect comparisons [Bucher] can be extended to risk ratios, risk differences, and continuous outcomes. Meta-regression and adjusted indirect comparisons [Bucher], however, are limited to trials with two treatment arms. Glenny and colleagues (13) support the use of random effects for either of these approaches. Analysts can readily use logistic regression, meta-regression, and adjusted indirect comparisons [Bucher] with standard statistical software such as SAS or Stata (see Glenny et al. Appendix 7). An advantage of logistic and meta-regression models is the possibility to adjust for potentially confounding factors, especially trial-level covariates such as length of follow-up, cointerventions, dosages, and others. Figure 4 summarizes considerations for different approaches graphically.

A disadvantage of all four approaches is that only trials with one common comparator can be included. Often, however, poor internal validity and heterogeneity severely limit the number of eligible trials.

More complex methods, such as network meta-analysis (23) or Bayesian methods (6;22), were not evaluated by Glenny and colleagues (13). Such approaches can handle less restrictive scenarios, but their use requires advanced statistical packages and expert statistical assistance. Network meta-analysis (23), for example, allows analysts to use direct evidence and accommodates multiple common comparators. With network meta-analysis, researchers may be able to use a greater overall number of trials, but depending on the effect of interest and complexity of "networks," the increase in power is not clear (23). Network meta-analysis attempts to estimate the inconsistency ("incoherence") for treatments effects with respect to evidence in the network.

HOW MANY STUDIES ARE NEEDED FOR INDIRECT COMPARISONS?

No guidance exists on how many studies are necessary to conduct indirect comparisons. If only few studies or only studies with low event rates are available, adjusted indirect comparisons can still render correct point estimates, but the precision of the estimates will be low. Essentially, they will lead to indeterminate results with wide confidence intervals. Confidence intervals of such findings often encompass clinically significant and nonsignificant differences, leading to indeterminate conclusions. Four times as many equally sized studies are necessary for an adjusted indirect approach to have the same power as a direct comparison (13). As stated above, analysts can address this issue in part by using more complex statistical approaches, such as network meta-analysis, that are not limited to one common comparator. However, these more complex methods require more assumptions, which are difficult, if not impossible, to verify, especially with a small number of studies, small sample sizes, or low number of events (i.e., small degrees of freedom).

https://doi.org/10.1017/S0266462308080240 Published online by Cambridge University Press

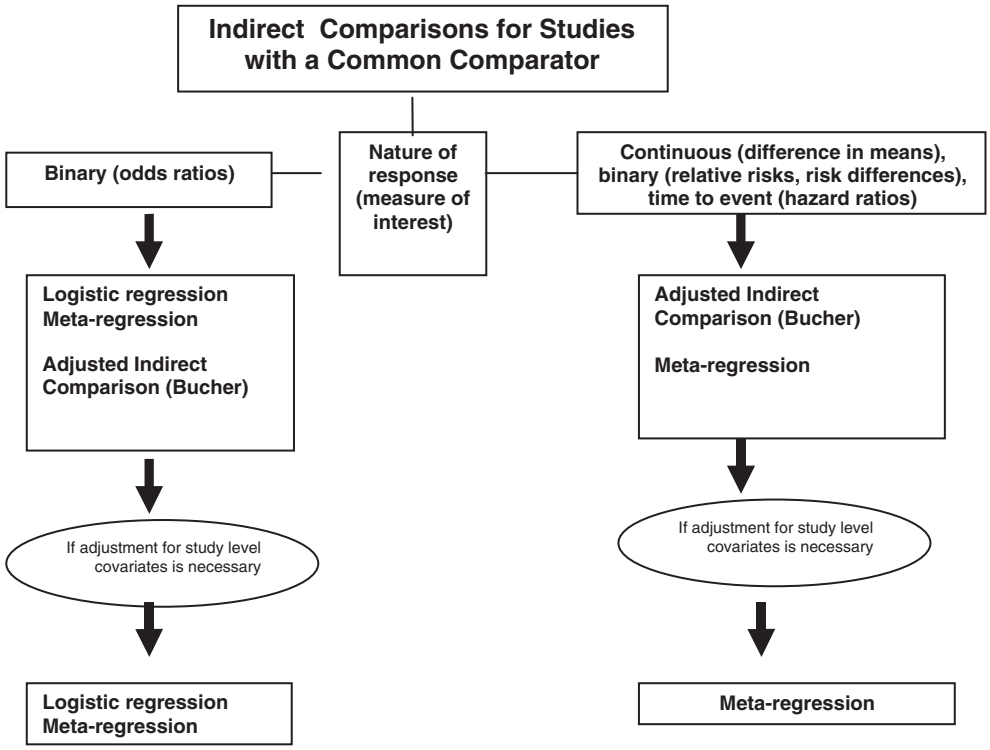


Figure 4. Algorithm for the choice of a statistical model for indirect comparisons.

CAN RESULTS FROM INDIRECT COMPARISONS AND HEAD-TO-HEAD TRIALS BE COMBINED?

Bucher and colleagues suggest that indirect comparisons should be conducted only when direct evidence is entirely missing (5). The Cochrane Collaboration handbook recommends that, in situations when both direct evidence and indirect evidence are available, the two approaches should be considered separately (16). They also recommend that, unless methodological issues compromise the head-to-head trials, direct evidence should take precedence as a basis for forming conclusions.

Glenny et al. (13) and Song et al. (39) take a more pragmatic approach, pointing out that sometimes direct randomized evidence is limited to few, often small studies with variable internal validity. Therefore, better, or at least more but equivalent information, or indeed information of higher quality might be available from indirect comparisons. Glenny and colleagues (13) recommend that analysts assess the combinability of studies by evaluating the similarity of study participants and interventions. They argue that statistical differences in direct and indirect estimates should not be the primary factor on which to make a decision whether to combine results. As in any RCTs, methodological issues (e.g., lack of allocation concealment, high differential loss to follow-up, and others) can compromise head-to-head RCTs and lead to biased results. Furthermore, head-to-head trials are often not powered to assess clinically relevant differences with enough statistical power to establish statistical significance (14;15). Particularly, head-to-head trials that were conducted primarily for marketing reasons often have small sample sizes. In such trials, even with sound randomization and allocation concealment, there can be substantial differences of prognostic factors between groups at baseline, resulting in selection bias.

Therefore, basing the decision on whether or not to combine direct and indirect estimates on the similarity or dissimilarity of effect estimates appears to be less valid than a careful examination of the similarity of study participants and interventions. If populations seem to be similar, then analysts can combine results from direct and indirect approaches, regardless of differences in effect estimates. Nevertheless, such an approach is based on theoretical considerations rather than on empirical evidence. Ultimately, deciding whether to combine direct and indirect estimates will be a matter of informed judgment. Statistically, results from direct and indirect comparisons can be easily combined using a weighted combination (inverse variance weighted average).

HOW SHOULD ANALYSTS HANDLE A SITUATION IN WHICH RESULTS FROM DIRECT AND INDIRECT COMPARISONS CONFLICT?

Glenny et al. (13) identified twenty-eight systematic reviews (with forty-four comparisons) in DARE in which enough data were available to conduct both direct and indirect

https://doi.org/10.1017/S0266462308080240 Published online by Cambridge University Press

comparisons. Overall, they observed a moderate agreement between direct and indirect comparisons (weighted kappa: 0.53). In three of forty-four comparisons, the differences between the direct and adjusted indirect comparisons reached statistical significance. Song et al. (40) compared results derived from direct and indirect comparisons based on "real world" systematic reviews. They reported considerable discrepancies between the two methods, but these could have been a result of the very small number of trials that were used in each systematic review (three to seven total trials).

No clear guidance yet exists on how to proceed when direct and indirect evidence conflicts. As mentioned above, evaluating the similarity of study participants and interventions appears to be the most constructive approach. Dissimilarities in study populations or interventions may explain conflicting results. If no dissimilarities exist, then results could be combined. However, presenting both combined and individual results may be the best and most informative approach.

To gain additional evidence on the comparative efficacy of sertraline versus venlafaxine, we conducted a meta-regression of placebo controlled studies. Results indicate that sertraline and venlafaxine have a similar antidepressive effect. The relative risk of achieving response is 0.88 (95 percent confidence interval, 0.72 to 1.07), which is consistent with the head-to-head trial (38) that indicated no difference in efficacy. Given the insufficient and conflicting data from head-to-head trials, these findings provide additional evidence to draw conclusions about the comparative benefits of these two antidepressants.

CONCLUSIONS

In the absence of sufficient head-to-head evidence, adjusted indirect comparisons can be viewed as an additional analytical tool to determine the comparative efficacy and effectiveness of competing interventions. Well-conducted methodological studies provide good evidence that adjusted indirect comparisons, when well done, can lead to results similar to those with direct comparisons. The internal validity of several statistical methods to conduct indirect comparisons, therefore, has been established. These methods studies, however, have been conducted under ideal circumstances. Researchers who consider indirect comparisons need to keep the limitations of indirect comparisons under "real-world" conditions in mind. First, given the limited information in many publications, the validity of indirect comparisons often has to be based on an unverifiable assumption. That assumption, essentially, requires that the effects of compared interventions be consistent across the different settings, populations, cointerventions, and other potentially confounding factors in the two sets of trials. Second, indirect comparisons are low in power, frequently leading to indeterminate results with wide confidence intervals. Inferences that can be drawn from such findings may be limited. Third, no consensus exists on how to interpret indirect results that differ substantially from direct evidence or on how to weigh findings of indirect comparisons against those results from nonrandomized direct evidence.

Many additional questions remain. At present, no consensus exists as to when conducting indirect comparisons or combining results from direct and indirect comparisons is appropriate. Furthermore, methods research has mainly focused on dichotomous outcomes, namely, to odds ratios. Future research needs to evaluate other binary outcome measures (e.g., risk ratios), continuous data, and time-to-event data. In addition, more complex statistical methods, such as network meta-analysis or Bayesian approaches that may increase power and precision by using multiple common comparators, have not been thoroughly validated. Future research also has to determine how indirect comparisons contrast with nonrandomized direct evidence in terms of accuracy of findings.

CONTACT INFORMATION

Gerald Gartlehner, MD, MPH (gerald.gartlehner@donau-uni.ac.at), Associate Professor, Department of Evidence-based Medicine and Clinical Epidemiology, Danube University, Karl Dorrek Straße 30, Krems, Austria, 3500
Charity G. Moore, PhD (moorecg@upmc.edu), Visiting Associate Professor, Department of Medicine, University of Pittsburgh, 200 Meyran Avenue, Suite 300, Pittsburgh, Pennsylvania 15238

REFERENCES

1. Alves C, Cachola I, Brandao J. Efficacy and tolerability of venlafaxine and fluoxetine in outpatients with major depression. *Primary Care Psychiatry*. 1999;5:57–63.
2. Baker SG, Kramer BS. The transitive fallacy for randomized trials: If A bests B and B bests C in separate trials, is A better than C? *BMC Med Res Methodol*. 2002;2:13.
3. Bennie EH, Mullin JM, Martindale JJ. A double-blind multicenter trial comparing sertraline and fluoxetine in outpatients with major depression. *J Clin Psychiatry*. 1995;56:229–237.
4. Boyer P, Danion JM, Bisserbe JC, et al. Clinical and economic comparison of sertraline and fluoxetine in the treatment of depression. A 6-month double-blind study in a primary-care setting in France. *Pharmacoeconomics*. 1998;13(pt 2):157–169.
5. Bucher HC, Guyatt GH, Griffith LE, et al. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. *J Clin Epidemiol*. 1997;50:683–691.
6. Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: Combining direct and indirect evidence. *BMJ*. 2005;331:897–900.
7. Costa e Silva J. Randomized, double-blind comparison of venlafaxine and fluoxetine in outpatients with major depression. *J Clin Psychiatry*. 1998;59:352–357.
8. Cunningham LA. Once-daily venlafaxine extended release (XR) and venlafaxine immediate release (IR) in outpatients with major depression. Venlafaxine XR 208 Study Group. *Ann Clin Psychiatry*. 1997;9:157–164.
9. De Nayer A, Geerts S, Ruelens L, et al. Venlafaxine compared with fluoxetine in outpatients with depression and concomitant anxiety. *Int J Neuropsychopharmacol*. 2002;5:115–120.

https://doi.org/10.1017/S0266462308080240 Published online by Cambridge University Press

10. Dierick M, Ravizza L, Realini R, et al. A double-blind comparison of venlafaxine and fluoxetine for treatment of major depression in outpatients. *Prog Neuropsychopharmacol Biol Psychiatry*. 1996;20:57–71.
11. Fava M, Hoog SL, Judge RA, et al. Acute efficacy of fluoxetine versus sertraline and paroxetine in major depressive disorder including effects of baseline insomnia. *J Clin Psychopharmacol*. 2002;22:137–147.
12. Fava M, Judge R, Hoog SL, et al. Fluoxetine versus sertraline and paroxetine in major depressive disorder: Changes in weight with long-term treatment. *J Clin Psychiatry*. 2000;61:863–867.
13. Glenny AM, Altman DG, Song F, et al. Indirect comparisons of competing interventions. *Health Technol Assess*. 2005;9:1–148.
14. Gotzsche PC. Lessons from and cautions about noninferiority and equivalence randomized trials. *JAMA*. 2006;295:1172–1174.
15. Greene WL, Concato J, Feinstein AR. Claims of equivalence in medical research: Are they supported by the evidence? *Ann Intern Med*. 2000;132:715–722.
16. Higgins JPT, Green S, eds. *Cochrane handbook for systematic reviews of interventions*. 4.2.5 [updated May 2005]. In: The Cochrane Library, Issue 3, 2005. Chichester, UK: John Wiley & Sons, Ltd.
17. Hypericum Depression Trial Study Group. Effect of Hypericum perforatum (St John's wort) in major depressive disorder: A randomized controlled trial. *JAMA*. 2002;287:1807–1814.
18. Ioannidis JP. Indirect comparisons: The mesh and mess of clinical trials. *Lancet*. 2006;368:1470–1472.
19. Khan A, Fabre LF, Rudolph R. Venlafaxine in depressed outpatients. *Psychopharmacol Bull*. 1991;27:141–144.
20. Kocsis JH, Zisook S, Davidson J, et al. Double-blind comparison of sertraline, imipramine, and placebo in the treatment of dysthymia: Psychosocial outcomes. *Am J Psychiatry*. 1997;154:390–395.
21. Kroenke K, West SL, Swindle R, et al. Similar effectiveness of paroxetine, fluoxetine, and sertraline in primary care: A randomized trial. *JAMA*. 2001;286:2947–2955.
22. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23:3105–3124.
23. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med*. 2002;21:2313–2324.
24. Lydiard RB, Stahl SM, Hertzman M, et al. A double-blind, placebo-controlled study comparing the effects of sertraline versus amitriptyline in the treatment of major depression. *J Clin Psychiatry*. 1997;58:484–491.
25. McDonagh MS CS. *Drug class review on proton pump inhibitors. Drug effectiveness review project 2006*. Available at: http://www.ohsu.edu/drugeffectiveness/reports/documents/PPIs%20Final%Report%u4%Unshaded.pdf. Accessed 11/2/2007.
26. Mehtonen OP, Sogaard J, Roponen P, et al. Randomized, double-blind comparison of venlafaxine and sertraline in outpatients with major depressive disorder. Venlafaxine 631 Study Group. *J Clin Psychiatry*. 2000;61:95–100.
27. Mendels J, Johnston R, Mattes J, et al. Efficacy and safety of b.i.d. doses of venlafaxine in a dose-response study. *Psychopharmacol Bull*. 1993;29:169–174.
28. Nemeroff CB, Thase ME. A double-blind, placebo-controlled comparison of venlafaxine and fluoxetine treatment in depressed outpatients. *J Psychiatr Res*. 2007;41:351–359.
29. Newhouse PA, Krishnan KR, Doraiswamy PM, et al. A double-blind comparison of sertraline and fluoxetine in depressed elderly outpatients. *J Clin Psychiatry*. 2000;61:559–568.
30. Olie JP, Gunn KP, Katz E. A double-blind placebo-controlled multicentre study of sertraline in the acute and continuation treatment of major depression. *Eur Psychiatry*. 1997;12:34–41.
31. Reimherr FW, Byerley WF, Ward MF, et al. Sertraline, a selective inhibitor of serotonin uptake, for the treatment of outpatients with major depressive disorder. *Psychopharmacol Bull*. 1988;24:200–205.
32. Reimherr FW, Chouinard G, Cohn CK, et al. Antidepressant efficacy of sertraline: A double-blind, placebo- and amitriptyline-controlled, multicenter comparison study in outpatients with major depression. *J Clin Psychiatry*. 1990;51(Suppl B):18–27.
33. Rudolph RL, Feiger AD. A double-blind, randomized, placebo-controlled trial of once-daily venlafaxine extended release (XR) and fluoxetine for the treatment of depression. *J Affect Disord*. 1999;56:171–181.
34. Schatzberg A, Roose S. A double-blind, placebo-controlled study of venlafaxine and fluoxetine in geriatric outpatients with major depression. *Am J Geriatr Psychiatry*. 2006;14:361–370.
35. Schneider LS, Nelson JC, Clary CM, et al. An 8-week multicenter, parallel-group, double-blind, placebo-controlled study of sertraline in elderly outpatients with major depression. *Am J Psychiatry*. 2003;160:1277–1285.
36. Sechter D, Troy S, Paternetti S, et al. A double-blind comparison of sertraline and fluoxetine in the treatment of major depressive episode in outpatients. *Eur Psychiatry*. 1999;14:41–48.
37. Silverstone PH, Ravindran A. Once-daily venlafaxine extended release (XR) compared with fluoxetine in outpatients with depression and anxiety. Venlafaxine XR 360 Study Group. *J Clin Psychiatry*. 1999;60:22–28.
38. Sir A, D'Souza RF, Uguz S, et al. Randomized trial of sertraline versus venlafaxine XR in major depression: Efficacy and discontinuation symptoms. *J Clin Psychiatry*. 2005;66:1312–1320.
39. 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.
40. Song F, Glenny AM, Altman DG. Indirect comparison in evaluating relative efficacy illustrated by antimicrobial prophylaxis in colorectal surgery. *Control Clin Trials*. 2000;21:488–897.
41. Thase ME. Efficacy and tolerability of once-daily venlafaxine extended release (XR) in outpatients with major depression. The Venlafaxine XR 209 Study Group. *J Clin Psychiatry*. 1997;58:393–398.
42. Tylee A, Beaumont G, Bowden MW, et al. A double-blind, randomized, 12-week comparison study of the safety and efficacy of venlafaxine and fluoxetine in moderate to severe major depression in general practice. *Primary Care Psychiatry*. 1997;3:51–58.
43. Tzanakaki M, Guazzelli M, Nimatoudis I, et al. Increased remission rates with venlafaxine compared with fluoxetine in hospitalized patients with major depression and melancholia. *Int Clin Psychopharmacol*. 2000;15:29–34.
44. Van Moffaert M, Bartholome F, Cosyns P, et al. A controlled comparison of sertraline and fluoxetine in acute and continuation treatment of major depression. *Hum Psychopharmacol*. 1995;10:393–405.

https://doi.org/10.1017/S0266462308080240 Published online by Cambridge University Press