

# Assessing exchangeability in indirect and mixed treatment comparisons

Demissie Alemayehu

Pfizer, Inc, Statistics Department,  
New York, USA

**Abstract:** In comparative effectiveness research (CER), investigators often resort to methods of indirect and mixed treatment comparisons, due to the unavailability of head-to-head comparative data from randomized clinical trials for competing treatment options. However, implicit in the available indirect comparison techniques is an assumption of exchangeability, which in practice cannot be conclusively verified. This paper discusses the implications of violations of this assumption, and describes approaches to evaluate its validity and steps that may be taken to minimize the impact on conclusions drawn from such studies.

**Keywords:** network meta-analysis, heterogeneity, comparative effectiveness research, systematic review

## Introduction

It is generally recognized that randomized controlled trials (RCTs) provide the most dependable estimates of the relative effects of alternative treatment options. In the hierarchy of evidence, a well planned and executed meta-analysis of data from RCTs holds a prominent place.<sup>1</sup> However, in many situations, head-to-head comparative data from RCTs are not available for all pair-wise comparisons that are of interest. Typically, RCTs intended for regulatory approval are conducted using a placebo as a control, so RCTs may not have been conducted comparing the new drug against standard treatments. In other situations, trials may exclude placebo use for ethical reasons, and there may be a need to quantify the true effect size of a new drug that has only been studied against an active control. In the context of comparative effectiveness research (CER), the lack of head-to-head comparative data creates additional challenges, since the scope of CER is wide and the focus is often on determination of the relative risk and benefits of all available treatment options.

To offset the problem arising from lack of data from head-to-head comparative RCTs, several indirect comparison procedures have recently been proposed in the literature.<sup>2-4</sup> The topic has also attracted the attention of regulatory agencies, in light of the significant implications of such analyses for health care decision-making.<sup>5,6</sup> In addition to the known limitations of traditional techniques used in systematic reviews,<sup>7</sup> the proposed indirect comparison techniques presuppose crucial assumptions that are peculiar to them. Some of the assumptions are inherent in the nature of the associated inference, which involves drawing conclusions about situations in which actual RCTs have not been conducted. Typically, techniques are constructed on the presupposition of similarity of relevant attributes and outcomes in trials. This poses considerable

Correspondence: Demissie Alemayehu  
Pfizer, Inc, 235 East 42nd, Street/219-8-57,  
New York, NY 10017, USA  
Tel +1 212 573 2084  
Fax +1 212 309 4873  
Email alemad@pfizer.com

challenges, since there is no definitive way of ruling out differences between trials in such important attributes as protocol requirements, temporal occurrences, patient characteristics and study personnel. The reliability of conclusions drawn from such analyses is therefore heavily dependent on the validity of these underlying assumptions.

## Indirect and mixed treatment comparisons

As an example, consider a simple case where the interest is in evaluating the comparative effect of two treatments, A and C, which have not been studied in a head-to-head RCT. Suppose, however, that the two treatments were each studied in separate RCTs relative to a common comparator B, with corresponding estimated effects,  $d_{AB}$  and  $d_{CB}$ . In application,  $d_{AB}$  and  $d_{CB}$  may denote suitable functions of the usual efficacy measures, including odds ratios, risk differences or mean differences. Bucher et al<sup>2</sup> proposed estimating the effect of A relative to C indirectly by

$$d_{AC} = d_{AB} - d_{CB}.$$

The indirect estimator, as opposed to the direct estimator based on the individual effect estimates of A and C, has the implicit advantage of preserving some aspects of the randomization. Also, if the original estimators  $d_{AB}$  and  $d_{CB}$  have desirable large-sample properties, so does the indirect estimator, thereby facilitating statistical inference (ie, construction of confidence intervals and performing tests of hypotheses).

The approach has been extended to handle more complicated situations, including when there is direct evidence from multiple sources, and when there is data both from direct and indirect comparisons. The network meta-analysis approach proposed by Lumley<sup>5</sup> is particularly useful when there is direct evidence from multiple paths that form a closed loop. An interesting aspect of this approach is that the model incorporates a term to assess the degree of agreement between effect estimates obtained through different paths, dubbed the “incoherence” of the network. More complex methods, commonly referred to as mixed treatment comparison (MTC), have also been developed to synthesize information both from direct and indirect sources.<sup>8,9</sup> Most of the MTC approaches are based on a Bayesian framework, and require a careful assessment of the consistency of evidence from the direct and indirect sources.

In all cases, the validity of the procedures depends on a number of crucial assumptions. Some of these assumptions

include those that are relevant even when one conducts traditional meta-analysis.<sup>7</sup> Others are peculiar to indirect comparison techniques.<sup>10,11</sup> The assumption that forms the primary focus of this paper is that of exchangeability. Loosely stated, exchangeability requires that the same effect size would be obtained if the A versus B comparison was performed under the conditions of the B versus C comparative trial, and vice versa. In other words, the relative efficacy of a treatment is the same in all trials included in the indirect comparison.

More formally, suppose  $E_{AB}$  denotes the experimental condition under which a RCT was conducted to compare treatments A and B. Let  $d_{AB}(E_{AB})$  denote the relative treatment effect of A versus B under experimental condition  $E_{AB}$ . Then exchangeability is satisfied if and only if:

$$d_{AB}(E_{AB}) = d_{AB}(E_{CB})$$

and

$$d_{CB}(E_{CB}) = d_{CB}(E_{AB}).$$

The above is, of course, a hypothetical construct, and cannot be tested using data. At a minimum, this would necessarily require both methodological similarities as well as comparability of patient populations. The former particularly presupposes correspondence with respect to the quality of the RCTs, the designs of the studies, patient and disease characteristics, and other experimental conditions, including study site and personnel.

The next section discusses a systematic approach to assessing the assumption of exchangeability. It considers the usual case, when only aggregate data is available, and also proposes approaches for situations when individual patient data may be accessible.

## Assessing exchangeability

In multicenter randomized controlled studies involving the comparison of A and B, efforts are generally made to ensure adherence to the research protocols. However, there are often treatment-by-center interactions, which might be a consequence of many factors, including the quality of the staff conducting the study.<sup>13</sup> While there are statistical techniques to assess the presence or absence of heterogeneity of treatment effects across study sites, there is no way to assess or quantify all the contributing factors, especially those relating to the quality of the study personnel at the sites.

In indirect comparisons, the problem is even more complex, since in the absence of data from RCTs comparing A

versus C, it is almost impossible to verify definitively the assumption of exchangeability. However, there are certain steps that may be taken to understand whether or not the assumption is tenable, and thereby to enhance confidence in the validity of the result.<sup>6</sup> Given the complexity of the issue, an effective framework is one that tackles the problem using a multipronged approach, including qualitative data, quantitative tools, and simulation techniques.

## Qualitative approaches

The first line of defense is to identify the factors that are known to “affect and therefore confound the comparative treatment effects.”<sup>6</sup> Such factors may consist of: study design parameters, including blinding; duration of treatment; efficacy measures; and dosing schedules. External factors, such as health care policy, location, and when the trials were conducted would also need to be evaluated. In addition, the patient populations studied should be compared with respect to relevant demographic characteristics, disease state, and medical history. Once the factors are identified, they should then be qualitatively assessed to ensure comparability, both between treatments within each study, and also across studies.

Like any observational study, indirect comparisons are likely to introduce bias emanating from the incomplete reporting of known covariates, as well as the impact of hidden or latent confounders. If the qualitative assessment suggests the presence of known confounders that have not been accounted for, appropriate measures should be taken to address the issue. When individual patient level data is available, suitable analyses should be performed incorporating the relevant factors into the model. For hidden or latent covariates, it may be appropriate to apply such techniques as instrumental variables developed for similar situations in nonrandomized studies.<sup>14,15</sup> When only aggregate data is available, metaregression is often recommended to adjust for the confounders.<sup>16</sup> However, the limitations of such an approach should be recognized, especially the bias resulting from the so-called ecological fallacy. The latter arises when patient-level covariates are aggregated at the study level and used in a regression analysis. Further, given the small number of studies that are typically included in indirect comparisons, there may not be adequate data for satisfactory model-building,<sup>17</sup> thus limiting the utility of the technique.

## Quantitative assessment

While the qualitative approach discussed above serves to give a general assessment of the degree of similarity of the

studies under consideration with regard to relevant factors, it needs to be complemented by quantitative analyses. When only aggregate data are available, the assessment may be limited to evaluation of the summary statistics. This may include evaluation of the consistency of effect estimates for the reference group across trials, with formal inference to rule out the effects of chance. If results of meta-analyses are used, one should also look for the absence of heterogeneity in the original analyses, taking into account the issues associated with the usual tests of homogeneity. Depending on the number of studies, one may also perform formal inference, as is commonly done in certain network meta-analysis formulations.<sup>8</sup>

When it is reasonable to assume that studies differ only with respect to known factors, and when individual patient data are available, a reasonable strategy may be to identify a subgroup of patients in each study who share relevant characteristics with the patient population in the other study. Once a subgroup is identified, then a test of homogeneity may be performed using this sub-grouping for stratification. The modified minimum volume ellipsoid (MVE) approach, which is extensively studied in the literature, is one such measure that may be applied to identify similar subgroups when the factors of interest are quantitative. More specifically, let  $X_{ij}$  denote the vectors of factors for the  $i$ th patient in the  $j$ th study. Define

$$d(X_{ij}, \bar{X}_R) = \sqrt{(X_{ij} - \bar{X}_R)^t S_R^{-1} (X_{ij} - \bar{X}_R)} \quad ,$$

where  $\bar{X}_R$  and  $S_R$  are robust estimators of location and covariance based on the target population in the study indexed by  $R$ . The idea is to identify those patients in the  $j$ th study that are close enough to the patients in  $R$ , as measured by  $d(X_{ij}, \bar{X}_R)$ , with respect to those characteristics quantified in  $X_{ij}$ .

While the MVE is a preferred approach, thanks to the robustness of the location and scale estimators against the influence of multivariate outliers, other distance measures, such as the Mahalanobis distance, could also be used. Alternatively, one could use a modified propensity score approach to identify similar groups. The modification involves performing the usual propensity score analysis, with study assignment as the dependent variable, rather than treatment, as is commonly done in such analyses. However, the utility of a propensity score is limited to known covariates.

Once similar subgroups are identified, a test of subgroup-by-treatment interaction may be performed, incorporating a term for the subgroup in the analytical model. If the analysis suggests that there is heterogeneity, then the assumption of

exchangeability cannot be justified, and appropriate remedial measures, as discussed above, should be taken.

## Simulation

Obviously, neither the qualitative nor the quantitative approaches discussed above are satisfactory in terms of giving unequivocal evidence for or against the tenability of the exchangeability assumption. In this regard, simulations may play an important role.<sup>18,19</sup> This typically involves developing predictive models based on an index trial, calibrating the model, and then simulating outcomes for patients in the other trials, based on the final model. For the simple scenario described in this paper, suppose a predictive model is constructed, including indicators for treatment, say A versus B, and other relevant predictors. The model is next calibrated for treatments B and C evaluated in the second study. The outcomes for patients in the second study are then simulated using the final calibrated model. Exchangeability may then be assessed by comparing the simulated and actual data according to a pre-specified criterion. This approach may be implemented both when aggregate data as well as individual patient data are available.

## Conclusion

Despite their routine use to establish the relative efficacy of drugs, RCTs are limited by their lack of external validity, and by the unavoidable absence of head-to-head comparative data on a range of treatment options. In this regard, indirect comparison procedures appear to offer a viable alternative option. However, in addition to the known problems associated with routine meta-analyses, such approaches depend on peculiar assumptions, one of which is that of exchangeability. Without this assumption, indirect comparisons cannot be performed, and at the same time it is almost always impossible to conclusively assess the validity of the assumption.

In this study a multipronged approach to evaluate the validity of the assumption of exchangeability is discussed. The framework involves a qualitative appraisal based on factors that are known to impact or modify treatment effects, as well as quantitative approaches using both aggregate and individual patient level data, when the latter is accessible. In addition, the role of simulation is highlighted to complement the qualitative and quantitative exercises.

It is recognized that the suggestions discussed above do not provide a definitive solution to the problem; however, they are intended to raise awareness about the impacts and complexity of the issue, and to suggest remedial measures. A major limitation of some of the approaches proposed is that

they require patient-level data for implementation. This, of course, is not plausible when information is being synthesized from aggregate data. Additional work is needed to explore other approaches to handle the issue when individual patient data may not be available. This paper also presents a general discussion of alternative strategies to address the exchangeability assumption in studies involving indirect and mixed treatment comparisons. It would be worthwhile to design a study or several case studies to validate the approaches. In addition, simulation experiments may be valuable to illustrate the performance of the assessment approaches as well as the remedial measures discussed in this paper. Further research is also needed to understand the operating characteristics of some of the indirect comparison techniques and to establish a solid theoretical foundation for assessing exchangeability as an integral part of the models.

With the growing interest in CER in the USA, there will be a corresponding dependence on results from indirect comparisons to make decisions with far-reaching consequences on public health and health care utilization. It will therefore be critically important to direct resources for further research to this area, and also to impose strict guidelines for the interpretation of results from such studies.

## Disclosure

The author is employed by Pfizer, Inc.

## References

1. Hierarchy of evidence and grading of recommendations. *Thorax*. 2004; 59 Suppl 1:i13–i14.
2. 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(6):683–691.
3. Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med*. 2004;23(20):3105–3124.
4. Hasselblad V. Meta-analysis of multitreatment studies. *Med Decis Making*. 1998;18(1):37–43.
5. Lumley T. Network meta-analysis for indirect treatment comparisons. *Stat Med*. 2002;21(16):2313–2324.
6. Wells GA, Sultan SA, Chen L, Khan M, Coyle D. *Indirect Evidence: Indirect Treatment Comparisons in Meta-Analysis*. Ottawa: Canadian Agency for Drugs and Technologies in Health; 2009.
7. ICWG. *Report of the Indirect Comparisons Working Group to the Pharmaceutical Benefits Advisory Committee: assessing indirect comparisons*. Available at: [http://www.health.gov.au/internet/main/publishing.nsf/Content/B11E8EF19B358E39CA25754B000A9C07/\\$File/ICWG%20Report%20FINAL2.pdf](http://www.health.gov.au/internet/main/publishing.nsf/Content/B11E8EF19B358E39CA25754B000A9C07/$File/ICWG%20Report%20FINAL2.pdf). Accessed May 31, 2011.
8. Egger M, Smith GD, Altman DG, editors. *Systematic reviews in health care: meta-analysis in context*. 2nd ed. London: BMJ Books; 2001.
9. Caldwell DM, Ades AE, Higgins JPT. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ*. 2005;331(7521):897–900.
10. Lu G, Ades AE. Assessing evidence inconsistency in mixed treatment comparisons. *J Am Stat Assoc*. 2006;101(474):447–459.
11. Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol*. 1986;15(3):413–419.

12. Song F, Altman DG, Glenny AM, Deeks JJ. Validity of indirect comparison for estimating efficacy of competing interventions: empirical evidence from published meta-analyses. *BMJ*. 2003;326(7387):472.
13. Adams JB, Pyke RE, Costa J, et al. A double-blind, placebo-controlled study of a CCK-B receptor antagonist, CI-988, in patients with generalized anxiety disorder. *J Clin Psychopharmacol*. 1995;15(6):428–434.
14. Greenland S. An Introduction to instrumental variables for epidemiologists. *Int J Epidemiol*. 2000;29(4):722–729.
15. Tannen RL, Wiener MG, Xie D. Replicated studies of two randomized trials of angiotensin-converting enzyme inhibitors: further empiric validation of the ‘prior event rate ratio’ to adjust for unmeasured confounding by indication. *Pharmacoepidemiol Drug Saf*. 2008;17(7):671–685.
16. Berlin JA, Santanna J, Schmid CH, Szczech LA, Feldman KA. Individual patient-versus group-level data meta-regressions for the investigation of treatment effect modifiers: ecological bias rears its ugly head. *Stat Med*. 2002;21(3):371–387.
17. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol*. 1996;49(12):1373–1379.
18. Caro JJ. Pharmacoeconomic analyses using discrete event simulation. *Pharmacoeconomics*. 2005;23(4):323–332.
19. Caro JJ, Möller J, Getsios D. Discrete event simulation: the preferred technique for health economic evaluations? *Value Health*. 2010;13(8):1056–1060.

### Comparative Effectiveness Research

### Publish your work in this journal

Comparative Effectiveness Research is an international, peer reviewed open access journal focusing on comparative effectiveness of health care including preventative health care strategies, diagnostic strategies, diagnostic technology, medical devices, drugs, medical technology, health systems and organization. The manuscript management system

Submit your manuscript here: <http://www.dovepress.com/comparative-effectiveness-research-journal>

Dovepress

is completely online and includes a very quick and fair peer-review system. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.