VALUE IN HEALTH 15 (2012) 940–947

Available online at www.sciencedirect.com

SciVerse ScienceDirect

journal homepage: www.elsevier.com/locate/jval

Comparative Effectiveness Research/Health Technology Assessment

Matching-Adjusted Indirect Comparisons: A New Tool for Timely Comparative Effectiveness Research

James E. Signorovitch, PhD[1,], Vanja Sikirica, PharmD, MPH[2], M. Haim Erder, PhD[2], Jipan Xie, MD, PhD[1], Mei Lu, MS[1], Paul S. Hodgkins, PhD, MSc[2], Keith A. Betts, PhD[1], Eric Q. Wu, PhD[1]*

[1]*Analysis Group, Inc., Boston, MA, USA;* [2]*Shire Development LLC, Wayne, PA, USA*

ABSTRACT

Objective: In the absence of head-to-head randomized trials, indirect comparisons of treatments across separate trials can be performed. However, these analyses may be biased by cross-trial differences in patient populations, sensitivity to modeling assumptions, and differences in the definitions of outcome measures. The objective of this study was to demonstrate how incorporating individual patient data (IPD) from trials of one treatment into indirect comparisons can address several limitations that arise in analyses based only on aggregate data. **Methods:** Matching-adjusted indirect comparisons (MAICs) use IPD from trials of one treatment to match baseline summary statistics reported from trials of another treatment. After matching, by using an approach similar to propensity score weighting, treatment outcomes are compared across balanced trial populations. This method is illustrated by reviewing published MAICs in different therapeutic areas. A novel analysis in attention deficit/hyperactivity disorder further demonstrates the applicability of the method. The strengths and limitations of MAICs are discussed in comparison to those of indirect comparisons that use only published aggregate data. **Results:** Example applications were selected to illustrate how indirect comparisons based only on aggregate data can be limited by cross-trial differences in patient populations, differences in the definitions of outcome measures, and sensitivity to modeling assumptions. The use of IPD and MAIC is shown to address these limitations in the selected examples by reducing or removing the observed cross-trial differences. An important assumption of MAIC, as in any comparison of nonrandomized treatment groups, is that there are no unobserved cross-trial differences that could confound the comparison of outcomes. **Conclusions:** Indirect treatment comparisons can be limited by cross-trial differences. By combining IPD with published aggregate data, MAIC can reduce observed cross-trial differences and provide decision makers with timely comparative evidence.

Keywords: comparative effectiveness, individual patient data, matching-adjusted indirect comparison.

Introduction

Health care decision makers face significant gaps between their needs for comparative effectiveness research (CER) and the limited availability of comparative data. The gap is particularly pronounced for new treatments, which are often integrated into treatment strategies and formulary policies without the benefits of randomized trials against all clinically or economically relevant alternatives. After the new treatment becomes available, observational studies based on registries or real-world data, or pragmatic trials, may be initiated. Such studies, however, will not provide reliable comparative evidence until sufficient outcomes data have accumulated. This delay in comparative evidence for new treatments reduces the value of CER for improving the decisions of physicians, payers, and patients.

An increasingly used approach for timely CER is the comparison of treatment outcomes across separate randomized trials. Detailed reviews of methodologies for such indirect comparisons have been published [1,2], and guidelines have been developed for researchers and decision makers [3–10]. By combining trials with overlapping comparator groups, multiple direct and indirect comparisons can be combined into a network meta-analysis that summarizes comparative evidence for all treatments in a therapeutic area. Although based on randomized trials, indirect comparisons and network meta-analyses involve comparisons of nonrandomized treatment groups and are akin to observational studies and subject to important limitations.

In particular, cross-trial differences in patients' baseline characteristics or differences in outcome definitions can bias indirect comparisons [1,2,11]. Although meta-regressions can adjust for cross-trial differences in baseline characteristics at an aggregate level, they cannot adjust for large numbers of baseline differences and may be subject to ecological bias [12,13]. A key assumption of indirect comparisons and network

We declare the following potential conflicts of interest: Vanja Sikirica, M. Haim Erder, and Paul S. Hodgkins are employees of Shire *Pharmaceuticals, Inc.*, and hold stock options in Shire *Pharmaceuticals, Inc.* James E. Signorovitch, Jipan Xie, Mei Lu, Keith A. Betts, and Eric Q. Wu are employees of Analysis Group, Inc., a consultant for Shire *Pharmaceuticals, Inc.*

* *Address correspondence to:* James E. Signorovitch, Analysis Group, Inc., 111 Huntington Avenue, 10th floor, Boston, MA 02199.
E-mail: jsignorovitch@analysisgroup.com.

1098-3015/$36.00 – see front matter Copyright © 2012, International Society for Pharmacoeconomics and Outcomes Research (ISPOR). Published by Elsevier Inc.
http://dx.doi.org/10.1016/j.jval.2012.05.004

meta-analyses is that cross-trial differences can be mitigated by measuring treatment effects relative to a common comparator (e.g., placebo). However, different modeling assumptions, embodied in the choice of a relative effect measure (e.g., relative risk, odds ratio, or risk difference), can lead to conflicting conclusions about comparative effectiveness. These limitations are difficult to address by using only published aggregate data, especially when only small numbers of trials are available.

In this article, we show that the use of individual patient data (IPD) from clinical trials for one treatment—but not necessarily all treatments—can address these limitations and that useful IPD are more readily available than currently appreciated. If IPD were available from all trials of interest, potential biases stemming from cross-trial differences could be mitigated by regression adjustment [14–18] or propensity scores [19,20]. Although IPD are seldom available for all trials, researchers engaged in CER can often access IPD for some trials. In particular, when CER is conducted or funded by a clinical trial sponsor, IPD could be available from the sponsor's trials. A recently developed statistical method—*matching-adjusted indirect comparison (MAIC)*—can combine IPD for some treatments with published summary data for comparator treatments. Through examples, we show how the incorporation of IPD into indirect comparisons via MAIC can 1) adjust for cross-trial differences in baseline characteristics, 2) reduce sensitivity to effect measures, 3) resolve differences in study outcome definitions, and 4) allow the comparison of clinically relevant dosages.

Methods

The MAIC approach has been published previously and is briefly reviewed below [21]. The approach can be applied in three steps.

Clinical trial selection

As in any data synthesis, a systematic review should be conducted to identify clinical trials for the treatments to be compared. Characteristics of the selected trials should then be carefully compared, including the study design (e.g., randomized or open-label trial), inclusion/exclusion criteria, baseline characteristics, outcome assessments (e.g., definitions of outcomes, schedule of assessments), and statistical methods (e.g., handling of early dropouts, baseline adjustment). Cross-trial differences in these features can be sources of heterogeneity in any meta-analysis. In indirect comparisons and network meta-analyses, these differences can also be sources of bias. The availability of IPD, which can provide opportunities to remove or reduce observed cross-trial differences, should be assessed for each trial.

Identification of outcome measures

A meaningful cross-trial comparison should focus on comparably defined outcome measures that are available in the included trials. The precise definition of the study outcomes, the schedule of assessments, the clinical relevance of different dosages, and the statistical methods used to summarize effects should all be considered. IPD should be reanalyzed to match the outcome definitions used in the published trial data as much as possible before making an indirect comparison. If outcome definitions cannot be matched exactly, sensitivity analyses should be considered.

Matching trial populations

In trials with IPD, patients who could not have enrolled in the published comparator trials (e.g., because of stricter inclusion/exclusion criteria) should be excluded from the indirect comparison analysis. Even after matching inclusion/exclusion criteria across trials, important cross-trial differences in patients' baseline characteristics can remain. To adjust for these differences by using MAIC, patients in trials with IPD are weighted such that their weighted mean baseline characteristics match those reported for the trials without IPD. This approach is a form of propensity score weighting in which patients in one treatment group (in this case the trial with IPD) are weighted by their inverse odds of being in that group versus the other treatment group (in this case the trial with only published aggregate data). The propensity score model can be estimated by using the generalized method of moments based on the aggregate data and IPD. Other baseline summary statistics such as medians and standard deviations can also be matched when available. Outcomes from common comparator arms (e.g., placebo) can be used to validate the matching process. After matching, continuous, binary, or time-to-event outcomes can be compared across balanced trial populations by using weighted statistical tests that incorporate the same weights developed in the matching process (e.g., using weighted *t* tests, weighted χ^2 tests, or Kaplan-Meier tests). Weighted statistical models (e.g., analysis of covariance) can also be used to ensure that similar methods are applied to all trials. Limitations of the MAIC approach are described in the Discussion section.

Example applications

To illustrate how the use of IPD with MAIC can address the limitations that arise for indirect comparisons without IPD, four example applications are presented.

Indirect comparisons with IPD can resolve significant differences in key baseline characteristics: vildagliptin versus sitagliptin in Japanese patients with type II diabetes mellitus [22]

Vildagliptin and sitagliptin are two treatments for type II diabetes that were recently approved for use in Japan. Both have been associated with better glycemic control compared with placebo or voglibose in randomized trials [23–29]. A systematic literature review of clinical trials in Japanese patients identified two vildagliptin [25,26] and two sitagliptin [23,27] trials. The common comparators included placebo and voglibose. An indirect comparison of aggregate data suggested that vildagliptin was associated with a significantly greater absolute decrease in mean % glycosylated hemoglobin A_{1c} (Hb A_{1c}) versus sitagliptin (difference = −0.17; 95% confidence interval [CI]: −0.33 to −0.01; $P = 0.024$) in the voglibose-controlled trials but that the difference was not statistically significant in the placebo-controlled trials (difference = −0.20; 95% CI: −0.45 to 0.05; $P = 0.133$). Significant cross-trial baseline differences in mean Hb A_{1c}, however, call into question the validity of the comparison based only on published aggregate data. Patients in the vildagliptin trials had significantly lower mean Hb A_{1c} at baseline than did patients in the sitagliptin trials. Higher baseline Hb A_{1c} has been associated with greater postbaseline Hb A_{1c} reduction in meta-analyses of oral antihyperglycemics [30]. Although meta-regression can adjust for baseline differences in aggregate trial data, it can be unreliable with data from only four trials.

By using IPD from the vildagliptin trials, patients were selected on the basis of inclusion/exclusion criteria specified in the sitagliptin trials and were reweighted to match exactly the baseline characteristics reported for the sitagliptin trials, including the baseline mean Hb A_{1c} as well as age, sex, body mass index, fasting plasma glucose (FPG), and diabetes duration. After matching, vildagliptin was associated with a significantly greater decrease in Hb A_{1c} compared with sitagliptin (Table 1). Compared with the indirect comparison based on only aggregate data, the treatment difference between vildagliptin and sitagliptin increased after MAIC, consistent with the expected effect of adjusting for baseline Hb A_{1c} differences. The use of IPD and MAIC in this example provided a more reliable comparison than did aggregate data alone,