VALUE IN HEALTH 22 (2019) 85–91

Available online at www.sciencedirect.com

ScienceDirect

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

Methodology

Alternative Weighting Approaches for Anchored Matching-Adjusted Indirect Comparisons via a Common Comparator

Helmut Petto, PhD [1,*], *Zbigniew Kadziola, MSc* [1], *Alan Brnabic, MSc* [2], *Daniel Saure, PhD* [3], *Mark Belger, BSc* [4]

[1]*Eli Lilly and Company, Vienna, Austria;* [2]*Eli Lilly and Company, Sidney, Australia;* [3]*Eli Lilly and Company, Bad Homburg, Germany;* [4]*Eli Lilly and Company, Erl Wood, UK*

ABSTRACT

Background: Adjusted indirect comparisons (anchored via a common comparator) are an integral part of health technology assessment. These methods are challenged when differences between studies exist, including inclusion/exclusion criteria, outcome definitions, patient characteristics, as well as ensuring the choice of a common comparator. **Objectives:** Matching-adjusted indirect comparison (MAIC) can address these challenges, but the appropriate application of MAICs is uncertain. Examples include whether to match between individual-level data and aggregate-level data studies separately for treatment arms or to combine the arms, which matching algorithm should be used, and whether to include the control treatment outcome and/or covariates present in individual-level data. **Results:** Results from seven matching approaches applied to a continuous outcome in six simulated scenarios demonstrated that when no effect modifiers were present, the matching methods were equivalent to the unmatched Bucher approach. When effect modifiers were present, matching methods (regardless of approach) outperformed the Bucher method. Matching on arms separately produced more precise estimates compared with matching on total moments, and for certain scenarios, matching including the control treatment outcome did not produce the expected effect size. The entropy balancing approach was used to determine whether there were any notable advantages over the method proposed by Signorovitch et al. When unmeasured effect modifiers were present, no approach was able to estimate the true treatment effect. **Conclusions:** Compared with the Bucher approach (no matching), the MAICs examined demonstrated more accurate estimates, but further research is required to understand these methods across an array of situations.

Keywords: Bucher method, entropy balancing, matching-adjusted indirect comparisons

Introduction

When a new medical intervention is introduced to the marketplace, health care payers, prescribers, and patients are interested in its efficacy and safety compared with currently available treatments. Because head-to-head randomized trials of a new treatment versus all competing treatments are not often available, adjusted indirect comparisons (AICs) as a form of network meta-analyses and meta-regression approaches [1] have become important components of health technology assessments (HTAs). Such methods are challenged by payers, who suggest that bias can be introduced into the treatment effects (TEs) when differences in the included studies exist and when there is an insufficient number of studies to perform a meta-regression. This includes differences between inclusion/exclusion criteria, definition of outcomes, and patient characteristics. Choice of an appropriate common comparator is also critical; that is, dosing and treatment regimens should be comparable across the studies. Matching-adjusted indirect comparisons (MAICs), introduced by Signorovitch et al. [2], have been proposed as a method to address these issues. MAIC is applicable when individual-level data (ILD) are available for at least one of the trials (or sets of trials) involved in the AIC. If ILD data are available, then for comparisons with aggregated data of other studies, there is an option to apply the

Conflicts of interest: The authors are employees of Eli Lilly and Company. Z. Kadziola, D. Saure, and A. Brnabic own shares in the company.

* *Address correspondence to:* Helmut Petto, Eli Lilly Regional Operations GmbH, Area Medical Center Vienna, Kölblg. 8-10, Vienna A-1030, Austria.

E-mail: petto_helmut@lilly.com

1098-3015/$36.00 - see front matter

https://doi.org/10.1016/j.jval.2018.06.018

same inclusion/exclusion criteria and use the same definitions of outcomes as in the study with available aggregated data. The MAIC approach then attempts to reduce bias in the AIC caused by remaining differences in patient populations between studies by matching ILD with published aggregate-level data (ALD) moments. This is not unique to MAIC because other proposed methods from the literature, including simulated treatment comparisons [3] and network meta-regression with limited ILD [4,5], attempt to adjust for TEs.

Signorovitch et al. [2] first proposed the use of an MAIC when ILD were available and provided a method for matching the ILD to the ALD by reweighting the outcomes of patients in the ILD. The general three-step approach suggested by Signorovitch et al. [6] for applying MAIC includes clinical trial selection, outcome measure identification, and trial population matching. To select trials, a systematic review is performed to identify publications related to the treatments for comparison. The availability of ILD in the selected trials is then assessed to provide opportunities to remove or reduce cross-trial differences. The identification of outcome measures involves a cross-trial comparison focusing on comparably defined outcome measures included in the selected trials. Before making an AIC, ILD should be reanalyzed to match the outcome definitions used in the published trial data if discrepancies are evident. Finally, it is important that trials with ILD consider excluding patients who could not have enrolled in the published comparator trials and adjusting for differences in baseline characteristics by weighting ILD, such that the weighted baseline characteristics match those in trials without ILD. A propensity score weighting approach can be used, where the propensity score model is estimated using the generalized method of moments on the basis of the ALD and ILD. The variance of the weighted TE can be assessed using the sandwich estimator approach or bootstrapping methods [7].

Attitudes toward MAIC methodology are changing, and these approaches are being accepted more readily by international HTA bodies for evidence synthesis [1,8–10]. Therefore, it is important to understand the limitations and assumptions of these methods when applied in practice. Issues that may affect the observed TEs include assessing the value of using the outcome from the common comparator arm in the matching process, understanding the role of unmeasured confounding or effect modifiers that may be available only in the ILD, investigating the effect of reweighting the ILD to mimic the imbalance observed in the ALD trials, and determining whether there are any additional benefits to using the entropy balancing weighting algorithm [11].

In this article, we explore the MAIC methodology on simulated data scenarios and specifically attempt to assess the following questions:

1. Is it better to perform the matching between ILD and ALD studies separately for active and control arms, rather than matching for active and control arms combined as suggested by Signorovitch et al. [2]?
2. Is there an advantage of including the outcome of a control treatment in the matching between ILD and ALD studies (in addition to common covariates)?
3. How biased are the results if unmeasured covariates are not accounted for in the matching?

Methods

Matching Methods and Indirect Comparisons

Available weighting methods attempt to match an ILD study to an ALD study (or studies) on the basis of common (baseline) covariates. The weighted outcomes for the treatments across the studies can then be compared. In this article, we focus on anchored AIC via a common comparator that is present in both studies.

To address our key questions, seven matching approaches (Table 1) were applied to six selected simulation scenarios.

No matching

The simplest method is if no weighting of the ILD study is performed. The results can be indirectly compared with the results from the ALD study using the Bucher et al. [12] method (Bucher in Table 1).

Matching-adjusted indirect comparison

With the MAIC as proposed by Signorovitch et al. [2], individual patient weights are chosen so that summaries of the covariate distribution (i.e., mean ± SD) of the weighted ILD study match those of the ALD study. The method is based on propensity score weights (equal to the odds of being in the ALD trial), where the weighted means of the ILD covariates exactly match those of the ALD trial. We primarily used the means for matching covariates (SigTotal in Table 1), but also investigated, as suggested by Signorovitch et al. [6], simultaneously matching on mean and variance (SigTotalVar in Table 1). Another adaptation of the method was proposed by Sikirica et al. [7]. The weights are based not only on covariates but also on matching on the outcome of the control treatment in the two studies. We used the mean outcome for this approach (SigOutcomeC in Table 1). Another adaptation of the MAIC that we implemented was to match the covariates separately for active and control arms in the two studies (SigArm in Table 1), and not combined as suggested in the original article of Signorovitch et al. [2].

For deriving AIC effect sizes, the Bucher et al. [12] approach could be applied using the weights derived by the different MAIC approaches and calculating weighted standard errors for the outcome in the ILD study that match. Nevertheless, this unrealistically assumes that the weights are known (not estimated). Therefore, a sandwich estimator [2] or a bootstrap approach [13,14] should be chosen. We applied the latter to achieve more accurate standard errors by calculating the weighted effect size in ILD per sample, similar to an approach in Sikirica et al. [7].

Entropy balancing

The entropy balancing method [11] is based on the method of moments estimation (moments of first, second, or higher order are exactly matched) and relies on a maximum entropy reweighting schema. Weights are chosen so that they are as close as possible to unit weights, penalizing extreme weighting schemes. As described for the MAIC approach proposed by Signorovitch et al. [2], we used means for matching on covariates (EbArm in Table 1). We extended the entropy balancing approach to balance not only on the ILD and ALD covariates but also on the active and control covariates for the ILD study (EbArmILD in Table 1). The same bootstrap approach used for the MAIC approach proposed by Signorovitch et al. was also used to obtain standard errors for the AIC.

Simulation Data Sets

Simulated data sets were created for two randomized controlled trials. A continuous outcome Y was compared in the ILD study between a treatment A and a control treatment C (the common comparator), and in the ALD study, between a treatment B and the same control treatment C. We were interested in the TE of Y from AIC A-C-B.

Simple scenarios (maximum two covariates) were used for detecting basic features of the weighting approaches. In 1000

Table 1 – Approaches for matching ILD to ALD and AIC methods applied

Matching approach (label and short description)	Details of matching approach (weighting method and variables matched on)	Method to estimate A-B AIC and its variance
Bucher (no matching and Bucher)	None	AIC and its variance were calculated using the Bucher et al. [12] method
SigTotal (Signorovitch match total)	Matching of ILD to ALD (regardless of treatment) was based on means baseline covariate X from each study using MAIC as proposed by Signorovitch et al. [2]	AIC = (weighted average of outcome in ILD arm A − weighted average of outcome in ILD arm C) − (average of outcome in ALD arm B − average of outcome in ALD arm C). Variance of AIC was based on the bootstrap approach
SigTotalVar (Signorovitch match by variance)	Matching of ILD to ALD (regardless of treatment) was based on mean and variance of baseline covariate X from each study using MAIC as proposed by Signorovitch et al. [2]	”
SigOutcomeC (Signorovitch match by outcome of control arm)	The matching between the ILD and ALD studies was based on common baseline covariate X (which is matched regardless of the treatments) and on the outcome Y of the control treatment. MAIC as proposed by Signorovitch et al. [2] was used for matching.	”
SigArm (Signorovitch match by treatment arm)	The matching between the ILD and ALD studies was based on baseline covariate X, but was performed separately for active treatments and for the control groups. MAIC as proposed by Signorovitch et al. [2] was used for the two separate matches.	”
EbArm (entropy balancing by treatment arm)	The matching of ILD to ALD data was based on baseline covariate X and was performed separately for active treatments and for the control groups. Entropy balancing was used for matching.	”
EbArmILD (entropy balancing by treatment arm of ILD study)	The matching between the ILD and ALD studies was based on common baseline covariate X and was performed separately for active treatments and for control groups. Simultaneously, a second baseline covariate Z (not available in the ALD study) was balanced between active treatment and control only in the ILD study. Entropy balancing was used for the matching.	”

AIC, adjusted indirect comparison; ALD, aggregate-level data; ILD, individual-level data; MAIC, matching-adjusted indirect comparison.

simulated data sets, each trial consisted of 300 patients. In addition, one simulated data set was created for a realistic ILD and ALD trial with six covariates (1000 patients in each). Standard errors of the AIC were derived from 1000 bootstrap resamples (for further details, see Supplemental Materials found at https://doi. org/10.1016/j.jval.2018.06.018).

The performance of each weighting method was assessed through the following (see Appendix Table 1 in Supplemental Materials found at https://doi. org/10.1016/j.jval.2018.06.018):

1. Estimated treatment effect $\hat{\Delta}$ and 95% confidence interval;
2. Bias, difference from “true treatment effect” Δ:

$$\text{Bias}(\hat{\Delta}) = \frac{\sum_{m=1}^{M}(\hat{\Delta}_m - \Delta)}{M};$$

3. Mean squared error:

$$\text{MSE}(\hat{\Delta}) = \frac{\Sigma_{m=1}^{M}(\hat{\Delta}_m - \Delta)^2}{M}.$$

Simulation Scenarios

There were six simple simulation scenarios in total, and these simple scenarios consisted of one or two normally distributed random baseline covariates (X, Z) in the ILD and ALD studies (X_{ILD}, Z_{ILD} and X_{ALD}, Z_{ALD}). The covariate X was available for analysis in both trials, whereas the covariate Z was available only in the ILD study.

The covariates were then related to one (continuous) outcome variable Y:

$$Y_{ILD} = f_{ILD}(T, X_{ILD}, Z_{ILD}) + \varepsilon_{ILD},$$

$$Y_{ALD} = f_{ALD}(T, X_{ALD}, Z_{ALD}) + \varepsilon_{ALD},$$

where f_{ILD} and f_{ALD} were linear models for getting mean outcomes, to which normally distributed error terms with mean 0 and variance 1 were added:

$$\varepsilon_{ILD},\ \varepsilon_{ALD} \sim N(0, 1)$$

The outcome Y was also influenced by the treatments (T) that were randomly assigned in the two studies. Variable T indicated either the control treatment C ($T = 0$) or the active treatments A or B ($T = 1$). If an interaction between treatments T and the covariate was included in the model, then covariate X or Z was an effect modifier. All six matching approaches (Table 1) were applied to the six simple simulated data sets.

The simple scenarios all had an expected effect size of 1.0. Scenario 0 included one baseline covariate X, which was not an effect modifier. Scenario 1 included one baseline covariate X, which was an effect modifier. Scenario 2 included two covariates X and Z; neither was an effect modifier, and Z was not included in the ALD study. Scenario 3 included two covariates X and Z; X was an effect modifier, and Z, although not an effect modifier, was not included in the ALD study. In scenario 4, both baseline covariates X and Z were effect modifiers, with Z not included in the ALD

study. Scenario 5 had one baseline covariate X, which was an effect modifier, and both the mean and variance differed between the ILD and ALD studies.

In scenario 6 we attempted to match under realistic study conditions. We assumed six baseline covariates (three binary and three continuous) differing between the two trials. In the studies ($j = 0$ for ILD, $j = 1$ for ALD) the outcomes Y_{Aj}, Y_{Bj}, and Y_{Cj} for treatment A, B, and C, respectively, were related linearly to the covariates X_{kj}:

$$Y_{Aj} = \alpha_j + \sum_{k=1}^{m} \beta_k X_{kj} + \varepsilon,$$

$$Y_{Bj} = \alpha_j + \alpha_B + \sum_{k=1}^{m} \left(\beta_k X_{kj} + \beta_{kB} X_{kj}\right) + \varepsilon,$$

$$Y_{Cj} = \alpha_j + \alpha_C + \sum_{k=1}^{m} \left(\beta_k X_{kj} + \beta_{kC} X_{kj}\right) + \varepsilon.$$

Here β_j indicates the study effect, β_B and β_C the TEs, and β_k the prognostic effect of covariates. Effect modifiers are indicated by interaction parameters β_{kB} and β_{kC}. The error terms were derived from normal distribution:.

$$\varepsilon \sim N(0, \sigma).$$

The outcome was simulated for all three treatments in both trials, so that the AIC A-C-B, from the matching approaches, could be verified by the ALD trial known difference. We also assumed that one covariate (Z) was not given in the ALD trial.

Results

We applied different matching approaches (Table 1) to our simulation scenarios 0 to 6. The results are shown in Figure 1 (for detailed numerical results, see Supplemental Materials found at https://doi. org/10.1016/j.jval.2018.06.018).

In scenario 0, in which the included covariate was not an effect modifier, all matching approaches resulted in nearly the same effect size for the AIC A-C-B as the unweighted Bucher method and were nearly identical to the expected A-B difference. The confidence intervals were slightly wider for Bucher, SigTotal, and SigTotalVar compared with the methods that matched separately on active and control treatment arms (SigArm, EbArm, and EbArmILD) and that also matched on the outcome of the control treatment (SigOutcomeC).

In scenario 1 we included one effect modifier. Methods matching the two studies separately for active and control treatments (SigArm, EbArm, and EbArmILD) produced the closest results to the expected A-B difference. Matching on the total (SigTotal and SigTotalVar) or on the outcome of the control (SigOutcomeC) performed slightly worse, and the weakest result was observed for the unmatched Bucher approach. Confidence intervals were in the same range for all methods.

In scenario 2 the two included covariates were not effect modifiers and one of them was not available for analysis in the ALD study. All methods performed well under this condition, except the one in which the MAIC approach was used to match also on the outcome of the control treatment (SigOutcomeC). Confidence intervals were in the same range for all methods.

In scenario 3 one of the two included covariates was an effect modifier and the other was not. Two approaches did not perform well under this condition. The unmatched Bucher approach (Bucher) and the MAIC approach for matching also on the outcome of the control treatment (SigOutcomeC) deviated from the expected A-B difference. Confidence intervals were in the same range for all methods.

In scenario 4 we included two effect modifiers, of which one was not available for analysis in the ALD. All matching methods failed to estimate the true A-B difference. The unmatched Bucher approach (Bucher) again performed the worst.

In scenario 5 the variance of one included effect modifier differed between the ILD and ALD studies. All matching methods, except the unmatched Bucher method (Bucher), came close to estimating the true A-B difference. Nevertheless, the MAIC approach, which was matching also on variances (SigTotalVar), gave a wider confidence interval.

In scenario 6 we simulated six realistic baseline covariates (effect modifiers and not effect modifiers). Under this condition, none of the matching approaches estimated the known treatment difference in the ALD study. The Bucher method overestimated the true effect, whereas the other methods underestimated it. The methods that matched the studies separately for active and control arms (SigArm, EbArm, and EbArmILD) performed the best. Confidence intervals were in the same range for all methods.

Discussion

In this article, we investigated different approaches to the matching of AICs, including no matching, the MAIC approach as proposed by Signorovitch et al. [2], and entropy balancing [11]. Our simulations and scenarios provide insight into how the selected matching methods perform. We focus here on AICs (anchored comparisons) as a common approach for HTAs.

The results demonstrated that where no effect modifiers are present (scenario 0) all methods perform equally well in estimating the TE. In the presence of one effect modifier, which is imbalance in ALD (scenario 1), the Bucher method and the matching on Signorovitch total methods or on the outcome of the control poorly estimated the true TE. When there are prognostic variables available for matching in the ILD but not available in the ALD (scenario 2), matching on the outcome of the common comparator introduced bias, which is not observed in the other weighting methods that matched on only baseline covariates. Given that unknown prognostic variables may not be commonly reported in the literature, we suggest this approach not be used. If prognostic variables are present in the ILD but not available in the ALD (scenario 3), forcing the balance within the ILD on these variables demonstrated a gain in precision. Again, the Bucher method and the matching on the outcome of the control performed poorly. Not surprisingly, when the unmeasured covariate was an effect modifier, all the matching methods introduced a bias (scenario 4).

When the effect modifier in the ILD and ALD had different variances (scenario 5), the Bucher method and the method matching on the variance did not accurately estimate the TE. The Bucher method performed poorly in estimating the effect, whereas the matching on variances slightly underestimated the TE but inflated the standard error of the TE. The original article by Signorovitch et al. [2] indicated that means of the covariates can be matched between ILD and ALD studies and that weights can be found that simultaneously match the variances. Incorporating the variances into the MAIC approach led to wider confidence intervals for the AIC. This extension is obviously adjusting correctly for the additional uncertainty. Thus, if different variances of an important effect modifier are observed, matching should be performed also on the variances and not only on means. In our scenario, the variance in the ALD was larger than in the ILD and we saw the benefits of such matching. Further work is required to investigate statistical properties of matching approaches if moments of higher order are affected.

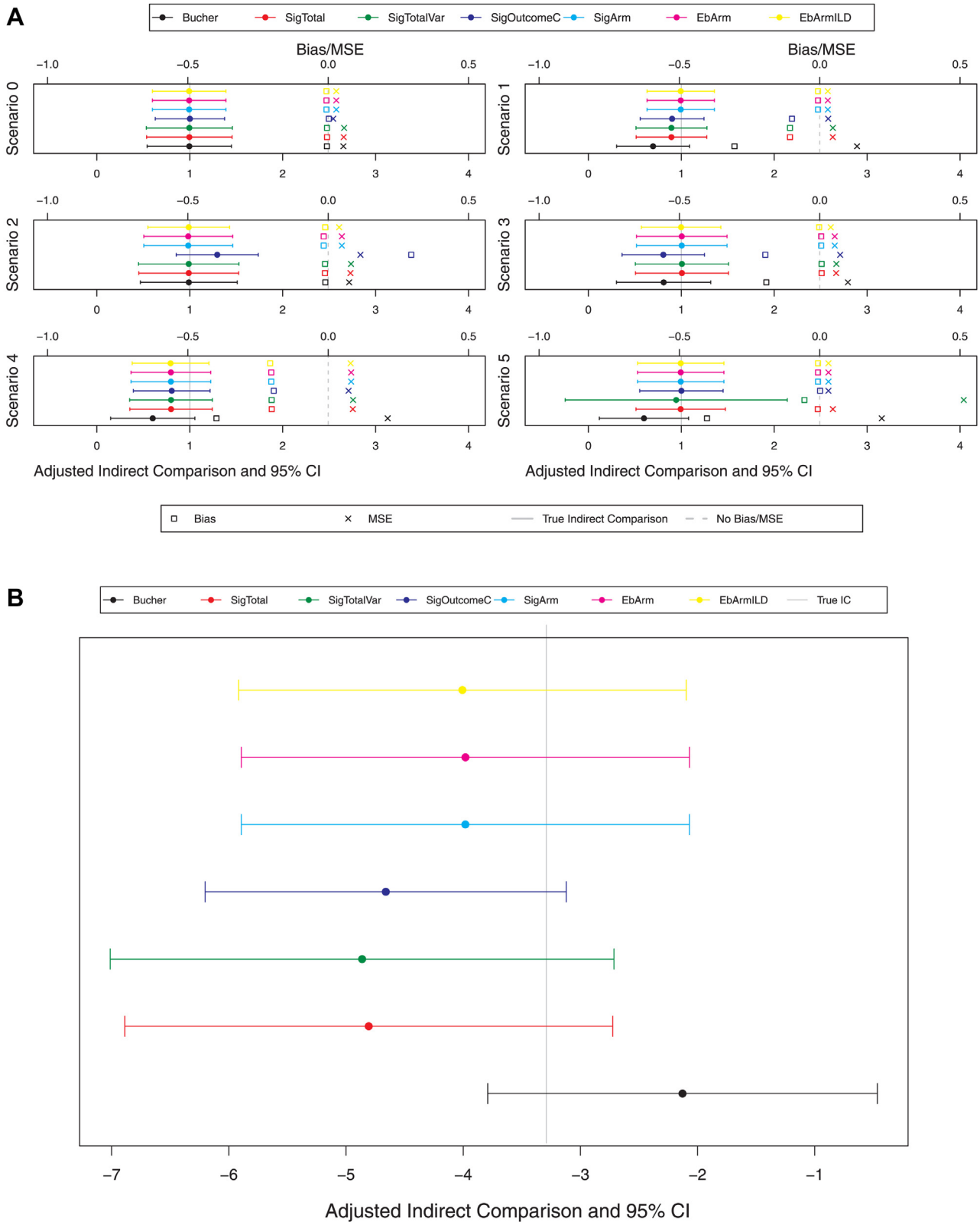

Fig. 1 – Forest plots for effect size of the indirect comparison A-C-B for scenarios 0 to 6. Indicated is the true expected effect size, the bias, and the MSE (detailed numbers given in Appendix Table 2 in Supplemental Materials found at https://doi.org/10.1016/j.jval.2018.06.018): (A) simple scenarios and (B) realistic scenarios. CI, confidence interval; MSE, mean squared error.

Our simulation scenarios have demonstrated that more precise estimates can be obtained when the matching is performed separately for active and control treatment arms between ILD and ALD studies versus matching on active and control arms combined as suggested by Signorovitch et al. [2]. When an imbalance existed between the ALD study arms (scenario 1), matching on the arms separately led in ILD to the same balance on observed covariates as in ALD and reduced the bias on the TE of A versus B. In randomized controlled trials, it is possible that unbalanced covariates of different magnitudes will be observed simply by chance. In such situations, the MAIC approach applied by treatment arm (matching approach SigArm) might be an alternative to the matching on the total population (matching approach SigTotal). Nevertheless, this alternative approach might break the randomization, and covariates might be imbalanced after weighting. Therefore, the choice of this approach should be carefully considered, and further investigation is needed to assess this approach under a more diverse array of situations.

Unbalanced effect modifiers are critical to determine whether a matching approach might give more accurate (unbiased) results than the simple Bucher approach. We therefore recommend checking whether effect modifiers might exist in the unbalanced covariates before the matching approach is selected. This check can be performed in the ILD study using regression techniques (e.g., including interaction terms between treatment and covariates), but cannot be performed for the ALD study. Therefore, additional approaches may be to consult an expert in this field or conduct a literature review to determine potential treatment modifiers that can be used in the analyses.

It is important to review the distribution of the weights when conducting these analyses because large weights will not only reduce the effective sample size but can also potentially lead to imbalances between treatment arms of the baseline covariates. We recommend checking the balance of covariates in the ILD and ALD trials between treatments before and after matching. If imbalances in either of the studies are detected after matching, then reweighting on individual treatment arms might be an alternative, but as already discussed, this approach has limitations.

Weights derived from the matching methods (the MAIC approach as proposed by Signorovitch et al. and entropy balancing) are directly related to the overlap of the covariates. The greater the distance, the more extreme the weights might be, so that a few data points might result in unrealistic large weights. So far, no formal decision rules of acceptable overlap have been established. The effective sample size [15] is based on the weights of the applied matching method. It may be used as a measure of how many patients were excluded because of the matching process, but it is not exact because it is assumed that the weights are known and not estimated.

For deriving the standard errors and thus the confidence intervals, we used in our simulations a bootstrap approach as suggested in the article by Sikirica et al. [7], which circumvented the assumption that weights were known. This approach preserved the variability arising from the estimation process. A statistical test then can be simply based on whether the resulting 95% confidence interval covered 0. As an alternative, the confidence intervals can be derived from sandwich estimators, as suggested by Signorovitch et al. [2].

There were several limitations in our approaches. First, in our simulation scenarios, we used a continuous outcome variable Y, but did not use a binary outcome. Further investigation is needed to identify the statistical properties of such an approach.

Second, in all the scenarios, a linear relationship between outcome and covariates was assumed. In medical research, many relevant outcomes are related at least approximately or in a relevant range linearly to covariates, but further investigation is required if the relationship is nonlinear, for instance, if outcome is time-to-event or if transformations are used to achieve linearity or if the relationship is different between ILD and ALD studies.

Third, the scenarios presented in this article may not reflect situations encountered in practice; we, however, attempted to overcome this by investigating the performance of the matching methods in a more realistic scenario (scenario 6). We assumed that five of six covariates were differing between the ILD and ALD trials and that one of the effect modifiers was not available in the ALD trial. The results showed that not only the Bucher method but also the more sophisticated matching approaches might give biased results. Matching the two studies separately for active and control treatment arms (SigArm, EbArm, and EbArmILD) gave better but still biased results. Finally, we did not investigate known limitations of unanchored direct comparisons [1] if only one treatment in the ILD was compared with one in the ALD trial.

Conclusions

As described, the MAIC approach is useful, in anchored settings, when effect modifiers are known and when marked imbalances exist between the ILD and ALD studies. At this stage, we consider the MAIC approaches presented here to be only sensitivity analyses used in HTA submissions. More research is needed before replacing the Bucher method [12] as the primary analysis for AIC. Alternatives that we have not examined are model-based approaches, such as simulated treatment comparison [3] and meta-regression for analyzing whole networks of studies, if at least for some ILD are available [4,5]. Further research is required to understand how the methods work when outcomes are either categorical or time-to-event; to understand what happens to the statistical properties of the methods if moments are of higher order; and to assess matching on the arms separately under more complex and varied scenarios.

Acknowledgment

This work was supported by Eli Lilly and Company. We thank Amber Burns, an employee of Eli Lilly and Company, for editing assistance.

Source of financial support: No funding was received for this study.

Supplementary Materials

Supplementary data associated with this article can be found in the online version at https://doi.org/10.1016/j.jval.2018.06.018.

REFERENCES

[1] Phillippo DM, Ades AE, Dias S, et al. NICE DSU Technical Support Document 18: Methods for population-adjusted indirect comparisons in submissions to NICE. 2016. Available from: http://www.nicedsu.org.uk. [Accessed January 30, 2017].

[2] Signorovitch JE, Wu EQ, Yu AP. Comparative effectiveness without head-to-head trials: a method for matching-adjusted indirect comparisons applied to psoriasis treatment with adalimumab or etanercept. Pharmacoeconomics 2010;28:935–45.

[3] Ishak KJ, Proskorovsky I, Benedict A. Simulation and matching-based approaches for indirect comparison of treatments. Pharmacoeconomics 2015;33:537–49.

[4] Saramago P, Sutton AJ, Cooper NJ, et al. Mixed treatment comparisons using aggregate and individual participant level data. Stat Med 2012;31:3516–36.
[5] Donegan S, Williamson P, D'Alessandro U, et al. Combining individual patient data and aggregate data in mixed treatment comparison meta-analysis: individual patient data may be beneficial if only for a subset of trials. Stat Med 2013;32:914–30.
[6] Signorovitch JE, Sikirica V, Erder MH, et al. Matching-adjusted indirect comparisons: a new tool for timely comparative effectiveness research. Value Health 2012;15:940–7.
[7] Sikirica V, Findling RL, Signorovitch J, et al. Comparative efficacy of guanfacine extended release versus atomoxetine for the treatment of attention-deficit/hyperactivity disorder in children and adolescents: applying matching-adjusted indirect comparison methodology. CNS Drugs 2013;27:943–53.
[8] National Institute for Health and Care Excellence. Dasatinib, nilotinib and imatinib for untreated chronic myeloid leukaemia, technology appraisal guidance. NICE Technology Appraisal Guidance TA426 2016. Available from: https://www.nice.org.uk/guidance/ta426. [Accessed January 30, 2017].
[9] National Institute for Health and Care Excellence. Bortezomib for induction therapy in multiple myeloma before high-dose chemotherapy and autologous stem cell transplantation, technology appraisal guidance. NICE Technology Appraisal Guidance TA311 2014. Available from: https://www.nice.org.uk/guidance/ta311. [Accessed January 30, 2017].
[10] Australian Government Department of Health, Pharmaceutical Benefits Advisory Committee. Guidelines for preparing submission to the Pharmaceutical Benefits Advisory Committee (PBAC), version 5.0. 2016. Available from: https://pbac.pbs.gov.au/content/information/files/pbac-guidelines-version-5.pdf. [Accessed January 30, 2017].
[11] Hainmueller J. Entropy balancing for causal effects: a multivariate reweighting method to produce balanced samples in observational studies. Polit Anal 2012;20:25–46.
[12] 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;50:683–91.
[13] Efron B. Bootstrap methods: another look at the jackknife. Ann Stat 1979;7:1–26.
[14] Efron B, Tibshirani RJ. An Introduction to the Bootstrap. New York, NY: Chapman & Hall/CRC; 1993.
[15] Kish L. Survey Sampling. New York, NY: Wiley; 1965.