


RESEARCH-IN-BRIEF

A critical assessment of matching-adjusted indirect comparisons in relation to target populations

Ziren Jiang¹ , Jialing Liu¹, Demissie Alemayehu², Joseph C. Cappelleri², Devin Abrahami³, Yong Chen^{4,5} and Haitao Chu^{1,2}

¹School of Public Health, Division of Biostatistics and Health Data Science, University of Minnesota, Minneapolis, MN, USA

²Statistical Research and Data Science Center, Pfizer Inc., New York, NY, USA

³Global Access and Value, Pfizer Inc., New York, NY, USA

⁴Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania, Philadelphia, PA, USA

⁵The Center for Health AI and Synthesis of Evidence (CHASE), University of Pennsylvania, Philadelphia, PA, USA

Corresponding author: Haitao Chu; Email: chux0051@umn.edu

Received: 5 April 2024; **Revised:** 29 January 2025; **Accepted:** 1 February 2025

Keywords: health technology assessment; indirect treatment comparison; MAIC paradox; matching-adjusted indirect comparison; population-adjusted indirect comparison

Abstract

Matching-adjusted indirect comparison (MAIC) has been increasingly applied in health technology assessments (HTA). By reweighting subjects from a trial with individual participant data (IPD) to match the summary statistics of covariates in another trial with aggregate data (AgD), MAIC enables a comparison of the interventions for the AgD trial population. However, when there are imbalances in effect modifiers with different magnitudes of modification across treatments, contradictory conclusions may arise if MAIC is performed with the IPD and AgD swapped between trials. This can lead to the “MAIC paradox,” where different entities reach opposing conclusions about which treatment is more effective, despite analyzing the same data. In this paper, we use synthetic data to illustrate this paradox and emphasize the importance of clearly defining the target population in HTA submissions. Additionally, we recommend making de-identified IPD available to HTA agencies, enabling further indirect comparisons that better reflect the overall population represented by both IPD and AgD trials, as well as other relevant target populations for policy decisions. This would help ensure more accurate and consistent assessments of comparative effectiveness.

Highlights

What is already known:

- Matching-adjusted indirect comparison (MAIC) methods are increasingly used in health technology assessment (HTA) submissions to adjust for population differences.
- MAIC estimates the comparative effectiveness of interventions for the population represented by the trial with aggregate data (AgD).

What is new:

- We present an illustration demonstrating an MAIC paradox in which the comparative effectiveness conclusions are reversed by switching the availability of IPD and AgD while adjusting the same set of effect modifiers.

- Additionally, we examine how variations in covariate distribution overlap of effect modifiers between trials influence the estimated comparative effectiveness.

Potential impact for Research Synthesis Methods readers:

- Through an illustrative example, we emphasize the vital importance of clearly defining the target population when applying MAIC in HTA submissions.
- We recommend providing de-identified IPD to HTA agencies to enable a more accurate assessment of comparative effectiveness for the target population.

1. Background of MAIC

Effect modification occurs when the magnitude of the effect of a treatment on an outcome differs depending on the value of a third variable. For example, studies indicate that Black individuals may experience less favorable outcomes compared to non-Black individuals when treated with angiotensin-converting enzyme (ACE) inhibitor-based therapies.^{1,2} In health technology assessments (HTAs), pharmaceutical companies are required to benchmark their new drugs against the prevailing standard of care for reimbursement decisions by HTA agencies.³ Nevertheless, the presence of effect modifiers can pose a unique challenge in comparing treatments when there is a lack of head-to-head trials. The traditional Bucher method,⁴ which compares the relative treatment effects of two interventions assessed in two randomized trials without covariate balancing, is limited to scenarios where all effect modifiers are balanced across trial populations.⁵ When individual participant data (IPD) are available for one trial while aggregate-level data (AgD) are only available for the other trial, researchers introduced population-adjusted indirect comparison (PAIC) methods to obtain unbiased estimates of comparative effectiveness, particularly when imbalances in effect modifiers exist. PAIC methods include matching-adjusted indirect comparison (MAIC),⁶ simulated treatment comparison (STC),⁷ and multilevel network meta-regression (ML-NMR).⁸

Among these methods, MAIC is becoming increasingly popular and widely used in health technology appraisals, such as the National Institute for Health and Care Excellence (NICE) in the United Kingdom.⁹ In a recent methodological systematic review, 88.9% (144 out of 162) of PAIC studies used MAICs.¹⁰ MAICs estimate a set of balancing weights for each subject in the IPD trial such that the weighted summary statistics (e.g., mean and standard deviation) of covariates in the IPD trial match the reported summaries of the same covariates in the AgD trial. Then, MAIC compares the marginal treatment effect estimated using the weighted data in the IPD trial with the marginal treatment effect reported in the AgD trial.^{5,11} For a more detailed description of the MAIC methods, readers can refer to the review paper by Jiang et al.¹²

Notably, the indirect comparison result is only valid with respect to the population represented by the AgD trial. However, this may not align with the population of interest for the company conducting the MAIC, which might prioritize the population represented by the IPD trial or another specific target population. This paper presents an illustrative example showing that MAICs can yield conflicting comparative effectiveness results when switching the availability of AgD and IPD between trials. This paradoxical phenomenon occurs due to differing magnitudes of effect modification for the two drugs by an effect modifier that is also imbalanced between the trial populations.

2. An illustrative example of the MAIC paradox

Consider an anchored indirect comparison between drug A (from Company A) and drug B (from Company B), each compared to a common placebo comparator C, as depicted in [Figure 1](#). Each company has access only to IPD from its own trial and to AgD for the other company's trial through published sources. For simplicity, we assume that race (Black versus non-Black) is the sole effect modifier, allowing both MAICs to include the "correct" effect-modifying variable. Additionally, we assume that drug A shows a stronger treatment effect among Black participants, while drug B is more

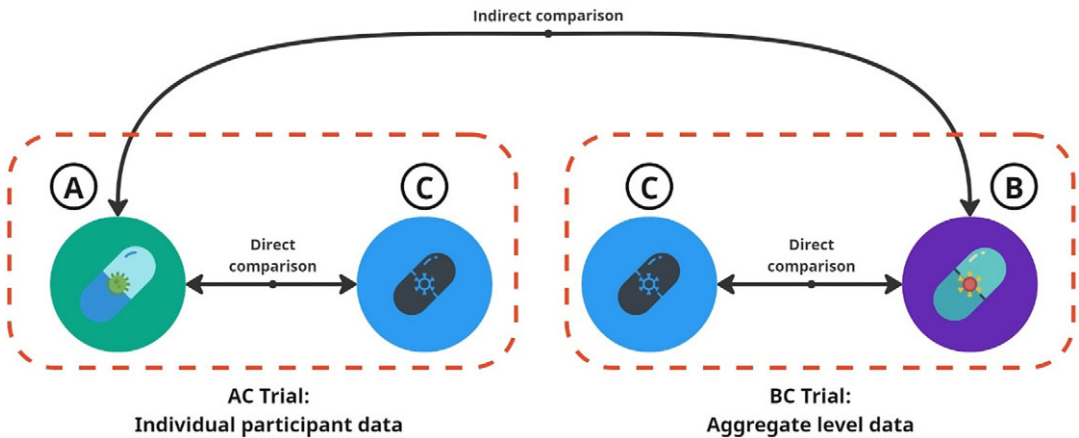


Figure 1. Indirect comparison of Drug A versus B in two trials. For the AC trial, we have the individual participant data (IPD). For the BC trial, we only have the aggregate level data (AgD).

effective among non-Black participants. If the AC trial includes a higher proportion of non-Black participants (among whom drug A is less effective than drug B), while the BC trial predominantly includes Black participants, we can observe a paradox: in separate MAICs, drug A outperforms drug B in the BC trial population, while drug B outperforms drug A in the AC trial population.

Table 1 presents hypothetical data of the AC and BC trials. Let $n_{11} = 800, n_{10} = 400$ be the number of non-Black and Black patients in the AC trial, and $n_{21} = 400, n_{20} = 800$ denote the number of non-Black and Black patients in the BC trial, respectively. The MAIC performed by Company A calculates the weights w_1 for all non-Black patients and w_0 for all Black patients in the AC trial such that the weighted proportion of non-Black patients $\frac{n_{11}w_1}{n_{11}w_1+n_{10}w_0}$ matches the proportion of non-Black patients in the BC trial $\frac{n_{21}}{n_{21}+n_{20}} = \frac{1}{3}$ (see Table 1), subject to the constraint that weights sum to 1 (i.e., $n_{11}w_1 + n_{10}w_0 = 1$). Solving the equation, we have $w_1 = \frac{1}{3n_{11}} = \frac{1}{2400}$ and $w_0 = \frac{2}{3n_{10}} = \frac{1}{600}$. Based on the trial data, under the usual logit link, the estimated treatment effect (log of the odds ratio of the survival rate) for drug A versus drug C in the population of BC trial would be

$$\log\left(\frac{\frac{80 \times w_1 + 180 \times w_0}{400 \times w_1 + 200 \times w_0}}{\left(1 - \left(\frac{80 \times w_1 + 180 \times w_0}{400 \times w_1 + 200 \times w_0}\right)\right)}\right) - \log\left(\frac{\frac{40 \times w_1 + 80 \times w_0}{400 \times w_1 + 200 \times w_0}}{\left(1 - \left(\frac{40 \times w_1 + 80 \times w_0}{400 \times w_1 + 200 \times w_0}\right)\right)}\right) = 1.540.$$

The indirect comparison of drug A versus B can then be obtained by subtracting the estimated marginal treatment effect for drug B in the BC trial from this value, which gives $1.540 - 1.115 = 0.425$. The corresponding standard error can be calculated using the robust sandwich estimator, which provides robust results given that the weights are estimated. The 95% confidence interval can then be constructed as (0.048, 0.802), indicating that drug A is statistically significantly better than drug B in the population of the BC trial.

Similarly, for the MAIC performed by Company B with IPD and AgD switched, the weights can also be determined by matching the weighted proportion of non-Black patients in the BC trial (IPD trial) to the proportion in the AC trial (AgD trial). This gives $w_1 = \frac{2}{3n_{21}} = \frac{1}{600}$ and $w_0 = \frac{1}{3n_{20}} = \frac{1}{2400}$. Performing the same calculations, the estimated comparative effectiveness for drug A versus drug B in the population of the AC trial is -0.402 with a 95% confidence interval $(-0.014, -0.790)$, which indicates that drug B is significantly better than drug A in the population of the AC trial.

Here, we emphasize that both conclusions—drug A being more effective than drug B or B being more effective than A—are potentially valid within the context of the specific populations considered

Table 1. Results for the illustrative example. In this example, the risk difference in survival rate for Drug A versus Drug C in the AC trial is 10% for non-black patients and 50% for black patients. Treatment effect for Drug B versus Drug C in the BC trial is 40% for non-black patients and 20% for black patients.

| | | AC Trial | | | BC Trial | | | A versus B |
|-----------|---------------|----------|--------|--------|----------|--------|-------|------------|
| | | Drug A | Drug C | logOR* | Drug B | Drug C | logOR | logOR |
| Non-black | Y = 0** | 80 | 40 | | 100 | 20 | | |
| | Y = 1 | 320 | 360 | | 100 | 180 | | |
| | n*** | 400 | 400 | | 200 | 200 | | |
| | Survival rate | 20% | 10% | 0.81 | 50% | 10% | 2.20 | -1.39 |
| Black | Y = 0 | 180 | 80 | | 240 | 160 | | |
| | Y = 1 | 20 | 120 | | 160 | 240 | | |
| | n | 200 | 200 | | 400 | 400 | | |
| | Survival rate | 90% | 40% | 2.60 | 60% | 40% | 0.81 | 1.79 |
| Overall | Survival rate | 43.3% | 20% | 1.12 | 56.7% | 30% | 1.12 | |

* logOR: Log of odds Ratio.

** Y: Outcome variable with Y = 1 indicating death and Y = 0 indicating survival.

*** n: Sample size.

in this example. However, without a clearly defined target population, the results of the indirect comparison lack meaningful applicability for guiding medical decisions.

In the supplementary material Section S1, we show a step-by-step derivation of the weights. In Section S2, we further explore the impact of varying proportions of non-Black participants in the IPD and AgD populations separately on the MAIC results (see Figures S1a and Figure S1b). In addition, we present an example illustrating this paradox in unanchored MAIC in Section S3.

3. Discussion

In this manuscript, we highlight a potential paradox, referred to as the “MAIC paradox,” where both companies may claim the superiority of their drugs through MAIC, even when the same covariates, including all effect modifiers, are included in the analyses. This paradox arises when there are imbalances in effect modifiers with different magnitudes of modification across treatments. The key issue is the lack of careful consideration of the target population. If the target population is not clearly defined or appropriately selected for the context, results from MAIC may lead to misleading or contradictory conclusions. This emphasizes the need for Health Technology Assessment (HTA) appraisals to explicitly define the target population that is most relevant for policy decision-making to ensure valid and consistent results. A clear definition of the target population is crucial to avoid misinterpretations that may arise from the MAIC paradox. Additionally, population overlap can impact MAIC results. Indirect comparison results tend to be more consistent when there is a high degree of overlap between populations. Conversely, if neither the AC trial population nor the BC trial population is comparable to the target population, MAIC may not reliably estimate the most relevant treatment effect. A more detailed discussion is provided in Supplementary Material Section S2.

The dependence of MAIC results on the target population has also been discussed in other literature as well. Following a review of NICE appraisals, Phillippo et al.⁹ observed that many appraisals overlooked the fact that comparative effectiveness was estimated over the population from the AgD trial, which might not represent the target population of interest. An example provided by the NICE DSU Technical Support Document 18¹¹ illustrates how contradicting conclusions may arise when both Novartis and AbbVie used MAIC to compare their drugs secukinumab and adalimumab: Novartis claimed significant efficacy advantages for secukinumab, while AbbVie argued that adalimumab had comparable efficacy but was more cost-effective. This discrepancy arose in part because two MAICs

used different sets of covariates in their analyses, which complicated the understanding of the cause for inconsistency. In our manuscript, we show that such paradoxes can still occur even when the correct MAIC model is used with the same set of covariates.

The manuscript illustrates how the target population affects the results of MAIC. However, the results of other PAIC methods, such as STC, also depend on the target population. Additionally, extrapolating outside the IPD population increases the risk of bias in the case of STC, which is a broader concern for any regression-based or other adjustment method. In the presence of effect modifiers, indirect comparison results are only valid when the target population is clearly defined.

Phillippo et al.⁵ proposed an additional sufficient condition for validly extrapolating the comparative effectiveness results to other populations, known as the “shared effect modifier assumption.” This assumption has two key components: 1) treatment effect modifiers are the same for all treatments, and 2) the magnitude of each effect modifier (i.e., how much it influences the treatment effect) is the same for all included treatments. Under the classical two-trial scenario, this assumption is statistically untestable, as only one trial provides IPD, leaving the other trial’s effect modifiers uncertain. Therefore, the validity of this assumption must only be demonstrated from a clinical perspective, relying on existing knowledge about the disease and treatments. This highlights the importance of ensuring clinical consistency and understanding of the treatments involved when applying MAIC or other PAIC methods.

Finally, we advocate for a collaborative effort among all relevant stakeholders to make de-identified IPD from clinical trials available through a trusted authority. Access to IPD from both trials would allow for a more robust approach to balancing the entire covariate distribution rather than just balancing the moments (such as means) of the covariates. This help mitigate the risk of ecological fallacy when drawing inferences between the two populations. However, sharing data must be done with the informed consent of trial participants, ensuring that appropriate de-identification protocols are followed to minimize the risk of participant re-identification.¹³ In addition to direct sharing of IPD, interim solutions could include the use of federated learning algorithms^{14–17} and secure data-sharing infrastructures,¹⁸ which allows for the analysis of summary statistics without exposing sensitive individual data.

Author contributions.

- Ziren Jiang: Methodology; Software; Writing - review & editing; Writing - original draft; Visualization.
- Jialing Liu: Writing - review & editing; Writing - original draft; Validation; Software.
- Demissie Alemayehu: Writing - review & editing; Writing - original draft; Validation; Supervision.
- Joseph C. Cappelleri: Writing - review & editing; Validation; Writing - original draft; Supervision.
- Devin Abrahami: Writing - review & editing; Writing - original draft; Validation; Supervision.
- Yong Chen: Supervision; Writing - review & editing; Writing - original draft; Validation.
- Haitao Chu: Conceptualization; Writing - original draft; Writing - review & editing; Methodology; Investigation; Supervision.

Competing interest statement. Demissie Alemayehu, Joseph C. Cappelleri, Devin Abrahami, and Haitao Chu are employed by Pfizer and own stocks in their company. However, all the contents in this manuscript are strictly educational, instructive, and methodological, not involving any real medicinal intervention.

Data availability statement. The simulated data used in the illustrative example are presented in [Table 1](#).

Funding statement. This work was supported in part by National Institutes of Health grants U01TR003709, U24MH136069, 1R01AG077820, R01AG073435, R56AG074604, R01LM013519, R01DK128237, R21AI167418 and R21EY034179 (Yong Chen).

Supplementary material. To view supplementary material for this article, please visit <http://doi.org/10.1017/rsm.2025.10>.

References

- [1] Exner DV. Lesser response to angiotensin-converting–enzyme inhibitor therapy in black as compared with white patients with left ventricular dysfunction. *N Engl J Med*. Published online 2001.
- [2] Ogedegbe G, Shah NR, Phillips C, et al. Comparative effectiveness of angiotensin-converting enzyme inhibitor-based treatment on cardiovascular outcomes in hypertensive blacks versus whites. *J Am Coll Cardiol*. 2015;66(11): 1224–1233. doi:[10.1016/j.jacc.2015.07.021](https://doi.org/10.1016/j.jacc.2015.07.021).

- [3] Dias S, Sutton AJ, Ades AE, Welton NJ. Evidence synthesis for decision making 2: a generalized linear modeling framework for pairwise and network meta-analysis of randomized controlled trials. *Med Decis Making*. 2013;33(5): 607–617.
- [4] 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.
- [5] Phillippo DM, Ades AE, Dias S, Palmer S, Abrams KR, Welton NJ. Methods for population-adjusted indirect comparisons in health technology appraisal. *Med Decis Making*. 2018;38(2): 200–211. doi:10.1177/0272989X17725740.
- [6] Signorovitch JE, Wu EQ, Yu AP, et al. 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(10): 935–945. doi:10.2165/11538370-000000000-00000.
- [7] Ishak KJ, Proskorovsky I, Benedict A. Simulation and matching-based approaches for indirect comparison of treatments. *PharmacoEconomics*. 2015;33(6): 537–549. doi:10.1007/s40273-015-0271-1.
- [8] Phillippo DM, Dias S, Ades AE, et al. Multilevel network meta-regression for population-adjusted treatment comparisons. *J R Stat Soc Ser A Stat Soc*. 2020;183(3): 1189–1210.
- [9] Phillippo DM, Dias S, Elsadat A, Ades AE, Welton NJ. Population adjustment methods for indirect comparisons: a review of National Institute for Health and Care Excellence technology appraisals. *Int J Technol Assess Health Care*. 2019;35(03): 221–228. doi:10.1017/S0266462319000333.
- [10] Truong B, Tran LT, Le TA, Pham TT, Vo T. Population adjusted-indirect comparisons in health technology assessment: a methodological systematic review. *Res Synth Methods*. 2023;14(5): 660–670. doi:10.1002/jrsm.1653.
- [11] Phillippo D, Ades T, Dias S, Palmer S, Abrams KR, Welton N. *NICE DSU Technical Support Document 18: Methods for Population-Adjusted Indirect Comparisons in Submissions to NICE*. Published online 2016.
- [12] Jiang Z, Cappelleri JC, Gamalo M, Chen Y, Thomas N, Chu H. A comprehensive review and shiny application on the matching-adjusted indirect comparison. *Res Synth Methods*. Published online February 21, 2024: jrsm.1709. doi:10.1002/jrsm.1709.
- [13] Ohmann C, Banzi R, Canham S, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open*. 2017;7(12): e018647. doi:10.1136/bmjopen-2017-018647.
- [14] Jordan MI, Lee JD, Yang Y. Communication-efficient distributed statistical inference. *J Am Stat Assoc*. 2019;114(526): 668–681. doi:10.1080/01621459.2018.1429274.
- [15] Luo C, Islam MdN, Sheils NE, et al. DLMM as a lossless one-shot algorithm for collaborative multi-site distributed linear mixed models. *Nat Commun*. 2022;13(1): 1678. doi:10.1038/s41467-022-29160-4.
- [16] Duan R, Ning Y, Chen Y. Heterogeneity-aware and communication-efficient distributed statistical inference. *Biometrika*. 2022;109(1): 67–83. doi:10.1093/biomet/asab007.
- [17] Duan R, Luo C, Schuemie MJ, et al. Learning from local to global: an efficient distributed algorithm for modeling time-to-event data. *J Am Med Inform Assoc*. 2020;27(7): 1028–1036. doi:10.1093/jamia/ocaa044.
- [18] Raisaro JL, Marino F, Troncoso-Pastoriza J, et al. SCOR: a secure international informatics infrastructure to investigate COVID-19. *J Am Med Inform Assoc*. 2020;27(11): 1721–1726. doi:10.1093/jamia/ocaa172.