The Effect Modification Problem: Using Stratified Matching Adjusted Indirect Comparison to Evade Misleading Interpretations of the Treatment Effect

Lawrence Rasouliyan, David Martínez

RTI Health Solutions, Barcelona, Spain

BACKGROUND

- Matching adjusted indirect comparison (MAIC) is a popular analytic method for estimating indirect treatment effects when one has access to individual patient data (IPD) from one trial and aggregate-level data (ALD) from another trial.[1]
- In situations where the IPD and ALD trials share a common comparator arm (anchored treatment comparisons), the National Institute for Health and Care Excellence (NICE) Decision Support Unit (DSU) guidelines recommend adjusting for all effect modifiers (EMs) to ensure balance between trials and to reduce bias in the treatment effect estimation.[2]
- As such, anchored MAIC ultimately estimates an overall indirect treatment effect in a specific population that has the distribution of EMs in the ALD trial. Thus, even with the use of identical methodologies, it is possible generate what appear to be conflicting results depending on which trial is IPD accessible versus ALD accessible due to the differences in the distributions of EMs across the trials.[2-3]
- This methodological shortcoming can greatly limit the generalisability of the findings. Additionally, it may be of interest to understand the estimated treatment effect for various subgroups.

OBJECTIVES

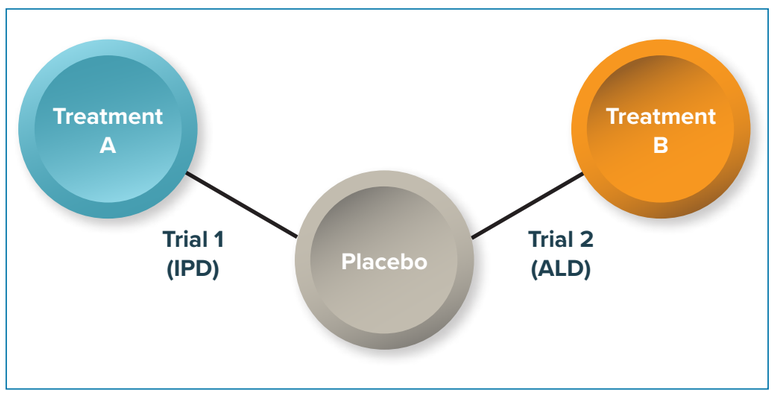
- Within the context of anchored MAIC, the objectives of this research are as follows:
  - To understand the influence of different EM scenarios on overall indirect treatment effect estimation
  - To estimate subgroup-specific indirect treatment effects for each level of each EM within the overall trial populations

METHODS

Data

- Patient-level example data sets were generated to simulate two oncology clinical trials, each with 1,000 patients (500 per arm) who met identical assumed entry criteria (Figure 1):
  - Trial 1: Treatment A vs. Placebo with accessible IPD
  - Trial 2: Treatment B vs. Placebo with accessible ALD

Figure 1. Schematic of Simulated Oncology Trial Network



- The primary endpoint was progression-free survival (PFS), which was simulated using the Weibull distribution.
- As is common in oncology trial publications, it was assumed that Trial 2 ALD reported subgroup analyses from which EMs were identifiable.
- The following dichotomous baseline characteristic variables were available in both trials and identified as EMs:
  - Genetic marker (EM in Trial 1 only)
  - Brain metastasis history (EM in Trial 2 only)
  - Late-stage disease (EM in both trials in the same direction)
  - Laboratory test positive (EM in both trials in opposite directions)
- Intervariable correlations between EMs in Trial 1 were assumed to reflect inherent biologic processes and were preserved in the simulation of Trial 2 data.

Statistical Methods

- Kaplan-Meier plots for PFS were generated for Trials 1 and 2 by treatment arm.
- Cox regression was employed in Trial 1 to compute the unadjusted hazard ratio (HR) and 95% confidence limits (CLs) of Treatment A vs. Placebo. This estimate was then combined with the HR of Treatment B vs. Placebo available in Trial 2 to compute a naïve HR of Treatment A vs. Treatment B (serving as a reference point for the unadjusted treatment comparison).
- Per NICE DSU recommendations, MAIC was implemented on overall trial populations adjusting for all EMs simultaneously:
  - Percentages of EM variables were tabulated for Trial 1 (pre- and postweighting) and Trial 2 to assess balance.
  - The overall adjusted Treatment A vs. Treatment B HR was calculated based on the EM distribution of the Trial 2 ALD.
- While applying IPD weights derived from the overall MAIC, stratified analyses were conducted:
  - For each EM, a separate Cox regression was performed in the Trial 1 data where the independent variables were treatment, the EM, and the treatment-EM interaction term.
  - These regressions resulted in adjusted Treatment A vs. Placebo HR estimates within each level of each EM.
  - Each EM-stratified estimate was combined with the analogous subgroup Treatment B vs. Placebo HR reported from Trial 2 ALD to compute EM-stratified adjusted Treatment A vs. Treatment B HR estimates.

RESULTS

- PFS Kaplan-Meier plots for treatment arms in Trials 1 and 2 are presented in Figure 2 along with corresponding HR estimates.
- Trial 1 patients were less likely to have the genetic marker and more likely to be laboratory positive compared with Trial 2 patients, while a similar distribution between trials was observed for the remaining EMs. After applying MAIC weights, EM variable percentages were almost identical between trials (Table 1).

Figure 2. Kaplan-Meier Plots for Trial 1 (IPD) and Trial 2 (ALD)

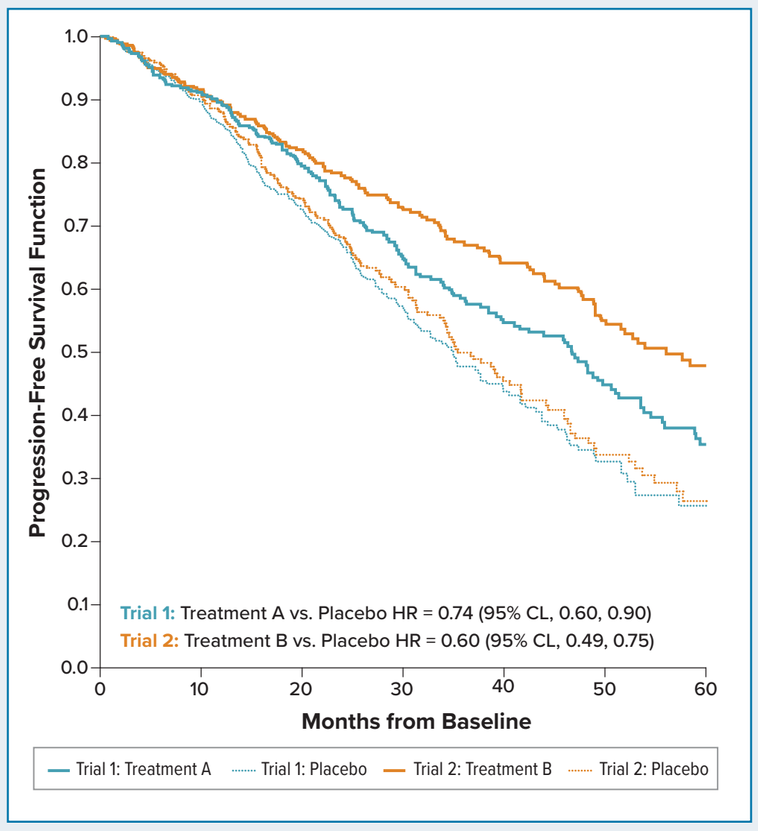


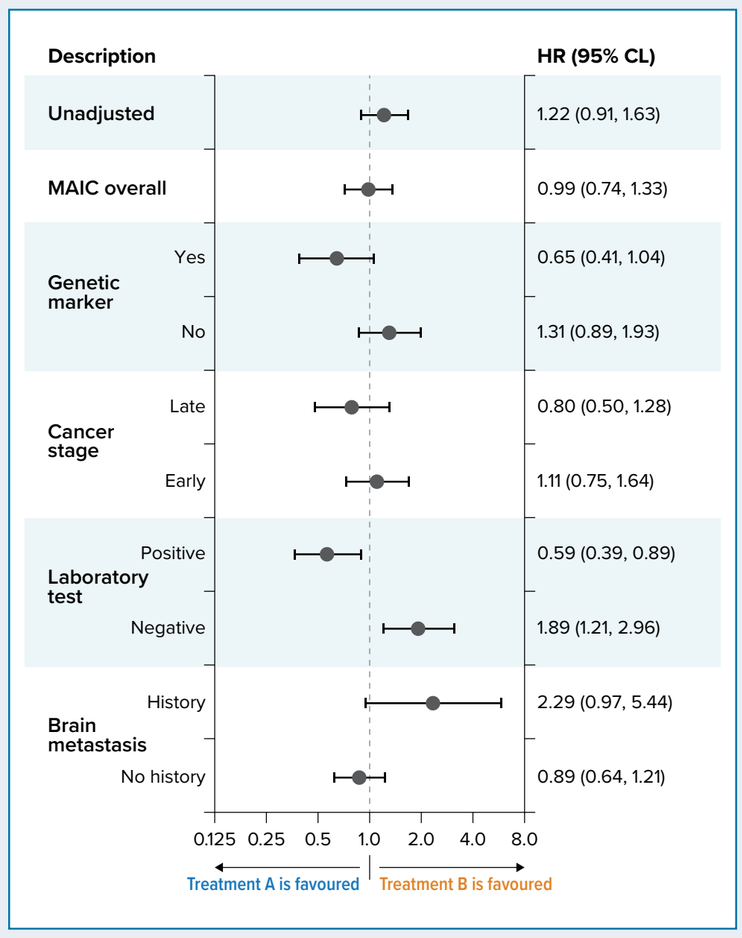
Table 1. Effect Modifier Distributions in Trials 1 and 2

Effect Modifier	Trial 1 (IPD) Preweighting N = 1,000	Trial 1 (IPD) Postweighting ESS = 756	Trial 2 (ALD) Reported N = 1,000
Genetic marker, %	25.2	38.7	38.3
Late-stage disease, %	41.6	44.1	44.1
Laboratory test positive, %	56.6	34.2	35.2
Brain metastasis history, %	14.4	14.7	14.7

ESS = effective sample size.

- Indirect comparisons of Treatment A vs. Treatment B are presented in Figure 3 as HR estimates and 95% CLs:
  - The unadjusted Treatment A vs. Treatment B HR was 1.22 (95% CL, 0.91, 1.63).
  - MAIC adjusting for all EMs simultaneously yielded an overall Treatment A vs. Treatment B HR of 0.991 (95% CL, 0.74, 1.33), suggesting an overall adjusted null treatment effect.
  - By design, Treatment A was favoured for laboratory test-positive patients (HR, 0.59; 95% CL, 0.39, 0.89), while Treatment B was favoured for laboratory test-negative patients (HR, 1.89; 95% CL, 1.21, 2.96).
  - MAIC EM-stratified HR estimates were notably different for patients with brain metastasis history (HR, 2.29) and patients without brain metastasis history (HR, 0.88), suggesting a preference for Treatment B over Treatment A in the former stratum but not in the latter.
  - Point estimates suggested a preference for Treatment A among patients with the genetic marker (HR, 0.65) and a preference for Treatment B among patients without the genetic marker (HR, 1.31).
  - A smaller amount of differential treatment effect was observed by cancer stage (late-stage HR, 0.80; early-stage HR, 1.11).

Figure 3. Treatment A vs. Treatment B HR Estimates



CONCLUSIONS

- Based on our simulation, the naïve (unadjusted) and overall MAIC analyses suggested a similar performance of Treatments A and B for PFS. This overall null treatment effect is the effect that would be observed in a population with the EM distribution of Trial 2.
- EM-stratified MAIC analyses demonstrated that Treatment A was favoured for some patient characteristics while Treatment B was favoured for others. Additionally, for some patient subgroups, the treatments performed similarly.
- In this simulation, the Treatment A vs. B comparison was truly different for various patient subgroups; however, these subgroup-specific effects cancelled each other when averaged across the entire population of Trial 2 to create an overall apparent null treatment effect.

DISCUSSION

- Consideration of reference population characteristics is imperative in the context of treatment effect interpretation.
- Because MAIC estimates the treatment effect in a specific reference population (that of the ALD), caution should be exercised when generalising findings to broader populations.
- Our recommendation is to report indirect estimates separately by EM category whenever possible, as it may help better inform policy decisions.
- These EM-stratified analyses assume that, after applying the overall MAIC weights to the IPD, cross-trial balance is maintained within each level of EM for the remaining EM variables. However, this assumption may be reasonable given likely inter-EM correlations due to inherent biologic processes.
- All MAIC analyses rely on information reported in the ALD, and the ability to perform EM-stratified analyses depend on the availability of subgroup analyses in the ALD.
- As with all indirect treatment comparisons, these statistical methods rely on many other important assumptions that cannot be tested directly.
- In practice, these methods cannot adjust for the potential effects of unmeasured EMs on the treatment comparison.
- More research is needed to understand the performance and implications of these methods in various scenarios of effect modification.

REFERENCES

1. Signorovitch JE, et al. Value Health. 2012 Sep-Oct;15(6):940-7.
2. Phillippo DM, et al. Med Decis Making. 2018 Feb;38(2):200-11.
3. Wang J, et al. Value Health. 2015;18(3):A31.

CONTACT INFORMATION

Lawrence Rasouliyan, MPH
Director, Biostatistics
Phone: +34.933.624.297
E-mail: lrasouliyan@rti.org

RTI Health Solutions
Av. Diagonal, 605, 9-1
08028 Barcelona, Spain

Presented at: ISPOR Europe 2019; 2-6 November 2019; Copenhagen, Denmark