Evaluating the Impact of Weighted Sample Size on Matching Adjusted Indirect Treatment Comparisons Between Trials with Time-to-Event Outcome: A Simulation Study

Ho-Yin Ho[1], Gabriel Tremblay*[1], Patrick Daniele[1]

1. Cytel, Inc., Waltham, MA, USA

Email: gabriel.tremblay@cytel.com

Background

- Matching-adjusted indirect comparisons (MAICs) play an essential role in supporting health technology assessment submissions by facilitating population-adjusted indirect treatment comparisons in the absence of head-to-head trials (Phillippo et al., 2016). MAICs rely on a propensity score (weighting) approach that rescales the weight of patients in the index trial to a target population. When estimating hazard ratios (HRs) between single-arm trials with time-to-event outcomes, an unanchored MAIC based on Cox regression can adjust for differences in measured baseline characteristics. In this kind of analysis, the sum of the MAIC weights of the index trial's individual patient-level data (IPD) may introduce varying degrees of bias and impact the efficiency of the estimated HR.

Objective: To explore the impact of sum of MAIC weights of IPD on bias and efficiency of the estimated HRs through a simulation study.

Methods

DATA GENERATION

- The parameters used to generate the simulation data are listed in Table 1. Age was not an effect modifier, but tumor grade was. The parameters in Table 1 are modified from a simulation study assessing MAIC performance by Jiang and Ni (2020).
- The sample size were set large enough to facilitate a reasonable performance of MAIC.
- Data were simulated for two single-arm studies from Weibull distributions under two scenarios:
  1. Tumor grade was an effect modifier. Parameters were set as shown in Table 1.
  2. Imbalance in tumor grade distributions between the two trials increased. Probability of tumor grade 1 was set to 10% instead of 40%.
- Data were simulated 2,000 times under each scenario.

Table 1. Data generation parameters

Parameter	Index trial	Comparator trial
Sample size	1000	800
Age	N(53,11^2)	N(56,12^2)
Range of age (truncation)	[21, 80]	[23, 85]
Tumor grade		
1	40%	50%
>=1	60%	50%
Scale parameter	0.00000004	0.00000004
Shape parameter	2.2	2.2
Prognostic effect		
Age	0.02	0.02
Tumor grade >=1	0.45	0.45
Effect modification		
Age	NA	0
Tumor grade >=1	NA	-0.2
Intercept	NA	-0.4
Treatment effect difference	NA	-0.5 - 0.2*50% = -0.5

SUM OF MAIC WEIGHTS

- The MAIC adjusted for both age and tumor grade. For the comparator trials, only aggregate summaries, mean (standard deviation) age and proportions of tumor grades, were used as inputs for MAIC.
- The analysis set w_i as the MAIC weight of the i-th patients of the IPD. Three types of sum of MAIC weights were considered:
  1. The raw weights: the direct outputs of MAIC without and with scaling.
  2. Effective sample size (ESS): the sum of MAIC weights of IPD was rescaled to be equal to the ESS. In other words, the sum of w_i was set equal to the ESS value.
  3. M1: the maximum MAIC weight of an individual observation was scaled to be 1. In other words, the scaled MAIC weight was w_i / $\max_i w_i$.
  4. Naïve: There was no MAIC weighting applied. All observations were weighted as 1.

ESTIMATION OF THE TREATMENT EFFECT

- Individual-patient time-to-event outcomes were assumed to be available for both index and comparator trials.
- Weighted Cox regressions only adjusting for the treatment were employed to estimate the treatment effect difference (HR).

RESULT EVALUATION

- The results of the simulation were summarized using statistical estimates of bias.
- Bias was computed using the difference between the estimated log HR minus the true value (-0.5) calculated using a naïve comparison and MAIC with difference sums of weights.

LIMITATIONS

- The conclusions of this study are limited by the simulation parameters and settings.
- For data with different structures, the conclusions, especially the ranking of the scaling strategies, are not expected to hold true.

Results

SCENARIO 1

- Figure 1 presents the 95% confidence interval (CI) of average bias in scenario 1. MAIC with any type of weighting demonstrated improvements relative to naïve comparison in terms of bias.
- The log HR estimator based on the M1 scaled MAIC weight was the least biased (0.0435, 95% CI 0.0414 to 0.0457), followed by the ESS scaled (0.0484, 95% CI 0.0463 to 0.0505), and raw (unscaled) (0.0489, 95% CI 0.0468 to 0.0510) MAIC weighting. The naïve estimator was the most biased estimator (0.0581, 95% CI 0.0560 to 0.0603).
- The average sum of MAIC weights with M1 scaled was 283 (72% reduction in sample size), which was much lower than with ESS (851; 15% reduction) or raw (930; 7% reduction) weighting.

SCENARIO 2

- Figure 2 presents the 95% CI of average bias. MAIC with any type of weighting remained unaffected compared to scenario 1. However, the naïve comparison suffered from noticeably larger bias.
- The log HR estimator based on the M1 scaled MAIC weight was the least biased (0.0381, 95% CI 0.0351 to 0.0412), followed by the ESS scaled (0.0414, 95% CI 0.0351 to 0.0412), and raw (unscaled) (0.0438, 95 CI% 0.04078 to 0.0468) MAIC weighting. The naïve estimator was the most biased estimator (-0.0871, 95% CI -0.0892 to -0.0849).
- The average sum of MAIC weights with M1 scaled (78.6) was again lower than ESS (319) and raw (567) weighting.
- The larger sample size reduction seen in scenario 2, compared to scenario 1, reflects the greater imbalance between trials.
- Overall, the results aligned with NICE DSU TSD18 (Phillippo et al., 2016), and MAIC was preferable to a naïve comparison when there were imbalanced effect modifiers.

Figure 1. Estimated 95% CI of bias under scenario 1

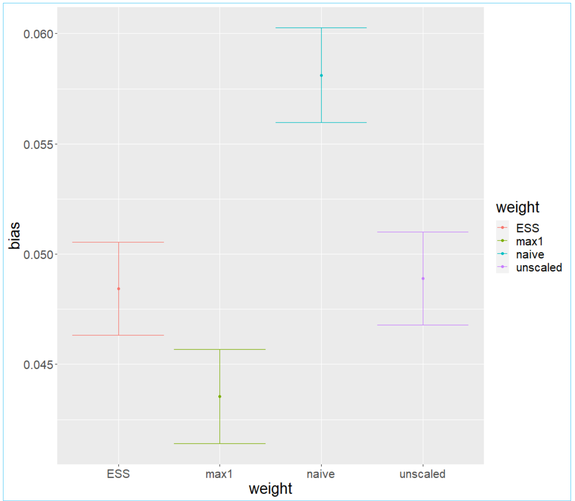
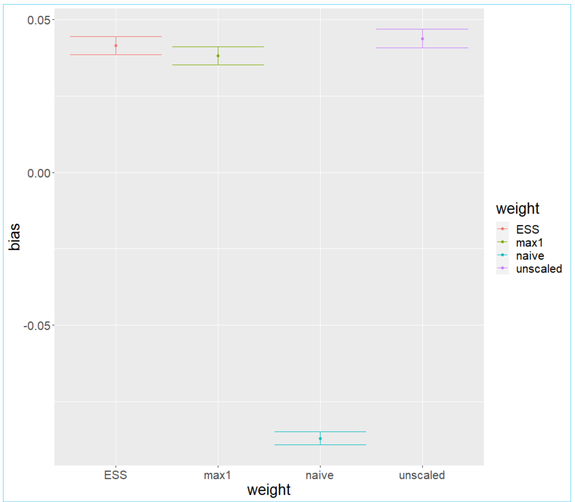


Figure 2. Estimated 95% CI of bias under scenario 2



Conclusions

- In indirect treatment comparisons, population matching is preferable to naïve comparisons when there are imbalanced effect modifiers. Prognostic variables are also important to consider when the comparison is unanchored.
- The choice of scaling of the sum of MAIC weights has a noticeable impact on bias and efficiency of the treatment effect difference estimator.
- Out of the scaling strategies considered in this study, M1 scaling led to the least biased estimator.
- Further analyses should be conducted to investigate the optimal scaling strategy for the potential gain in estimation efficiency.

Abbreviations: CI: confidence interval; ESS: effective sample size; HR: hazard ratio; IPD: individual patient-level data; MAIC: Matching-adjusted indirect comparison

References

1. Jiang, Y., Ni, W. Performance of unanchored matching-adjusted indirect comparison (MAIC) for the evidence synthesis of single-arm trials with time-to-event outcomes. *BMC Med Res Methodol* **20,** 241 (2020). https://doi.org/10.1186/s12874-020-01124-6
2. Phillippo, D. M., Ades, A. E., Dias, S., Palmer, S., Abrams, K. R., & Welton, N. J. (2016). NICE DSU technical support document 18: methods for population-adjusted indirect comparisons in submissions to NICE. *Report by the Decision Support Unit*