Evaluating the impact of random forest on matching-adjusted indirect comparisons of treatments between trials: A simulation study

Hoora Moradian[1], Bart Heeg[1], Gabriel Tremblay[3]
[1]Cytel, Inc., Waltham, MA, US

Background

- Matching-adjusted indirect comparison (MAIC) is a popular method of population-adjusted indirect treatment comparison. It uses a propensity score approach, which re-weights patient characteristics in the index trial to match the baseline characteristics of a target population.
- Compared with simulated treatment comparison, MAIC has the advantage of providing marginal relative treatment effect, whereas simulated treatment comparison only provides conditional relative treatment effect.
- The primary challenge with MAICs is that re-weighting may produce significantly smaller effective sample sizes (ESS; interquartile range: 48.0% to 84.6%)[1] which worsens and limits the results particularly when there is a high imbalance between the matching covariates of the two trials/studies and/or when the index trial has a small sample size.
- Random forest[2] (RF) is a non-parametric ensemble technique that averages outcomes from multiple decision trees and can be used to weight patient characteristics based on the number of times any pair of subjects ends up in the same terminal nodes.

Objective

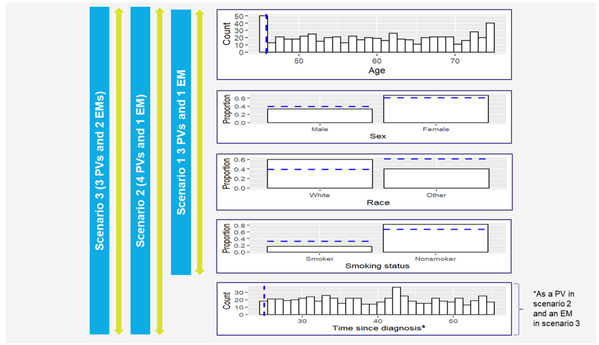
- The objective of this study was to explore the impact of using RF as the weight calculation technique compared with a propensity score approach on the results of MAICs. The impact of a high level of imbalance in effect modifiers (EM) between studies was evaluated in different scenarios which were varied based on the sample size and number of covariates.
- The hypothesis was that the RF approach estimates the weights more accurately vs the propensity score approach particularly when the latter's performance deteriorates, i.e., when there is high imbalance between the matching covariates of the two trials/studies and/or when the index trial has a small sample size, thereby improving the results of MAICs.

Methods

- The wakefield[3] package in R 4.2.0[4] was used to simulate data for a two-study comparison with logit link function with three treatment levels in three scenarios differing in terms of number of variables. Scenario 1 included three covariates, an EM (age) and three prognostic variables (sex, race, and smoking status). Scenario 2 included an additional prognostic variable (time since diagnosis) compared with scenario 1. Scenario 3 included time since diagnosis as an effect modifier instead of as a prognostic variable compared with scenario 2.
- In all scenarios, weights were estimated to match the EM distributions between the two trials using the RF and propensity score approach and MAICs were applied over 1,000 iterations. All scenarios were explored at three different sample size values for the index trial (N=50, 150, and 300). The randomForestSRC[5] package was used to calculate the RF weights and the maic[6] package was employed for MAIC analyses.
- Figure 1 illustrates the distributions of the variables of the index trial (AB) where there are individual patient data (IPD), compared with the reported average in a comparator trial (AC) where there are only summary data.

Methods (cont.) and Results

Figure 1. Variables distribution of the AB trial compared with the reported aggregate summary in the AC trial shown by dashed blue line in scenarios 1 to 3



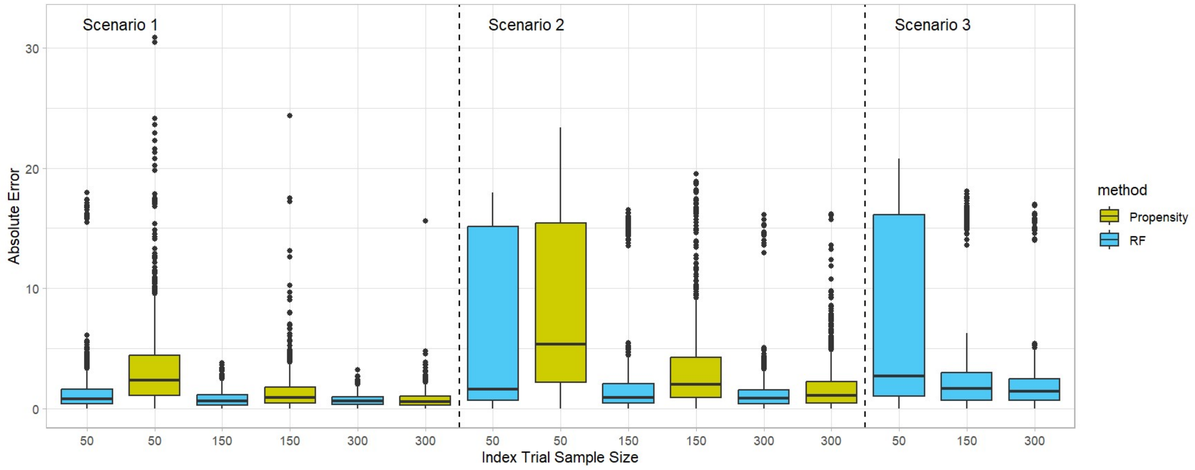
Abbreviations: EM, effect modifier; PV, prognostic variable

Figure 2. ESS reduction on average, mean absolute error for the true log odds ratio, number of times (out of 1,000 iterations) MAIC did not converge

Scenarios	Scenario 1			Scenario 2			Scenario 3		
Sample Size	50	150	300	50	150	300	50	150	300
RF ESS Reduction (%)	45.42%	68.42%	71.29%	31.46%	65.37%	73.36%	28.51%	61.49%	70.43%
Propensity ESS Reduction (%)	86.90%	86.77%	86.64%	87.04%	86.76%	86.80%	NA	NA	NA
RF Mean Absolute Error	1.42	0.80	0.70	5.91	2.28	1.29	6.93	3.02	1.88
Propensity Mean Absolute Error	3.44	1.38	0.77	7.75	3.29	1.72	NA	NA	NA
Number of Non-convergence Runs with RF	0	0	0	0	0	0	0	0	0
Number of Non-convergence Runs with Propensity	12	0	0	22	0	0	1,000	1,000	1,000

Abbreviations: ESS, effective sample size; NA, not applicable; RM, random forest

Figure 3. Absolute error estimating the true log odds ratio in MAICs with weights calculated with RF compared with the propensity score technique over 1,000 iterations



Note: MAICs using the propensity score weighting approach had no results for scenario 3 since it failed to converge in all iterations.
Abbreviation: RF, random forest

Results

- When the level of imbalance between EMs across the two trials was high, MAICs using the RF weighting approach had significantly lower mean absolute error as well as lower ESS reduction percentage compared with the propensity score weighting approach in scenarios 1 and 2 at all values of sample sizes (all p<0.001); the superiority increased as the sample size decreased from 300 to 50. (Figure 2)
- In scenario 3, MAICs using the RF weighting not only converged in all 1,000 runs, but also resulted in mean absolute error and ESS reduction (%) values similar to corresponding values of sample size in scenario 2. (Figure 2)
- Figure 3 shows the absolute error estimating the true log odds ratio in MAICs with weights calculated with RF compared with propensity technique over 1,000 iterations at a high level of imbalance.
- MAICs using the propensity score weighting approach failed to converge.

Conclusions

- The simulation studies supported the hypothesis that MAICs using the RF weighting approach provides a promising alternative to MAICs with propensity score weighting approach when there is a high imbalance between the matching covariates of the two trials/studies and/or when the index trial has a small sample size.
- MAIC with RF did not face the non-convergence issues common with the propensity approach when there is a high imbalance between the index trial and the comparator study and/or the index trial has a small sample size. In addition, it resulted in significantly higher accuracy and smaller percentage of ESS reduction.
- Further simulation studies should be conducted to explore results with other methods like MAIC with exact logistic regression[7] and parametric g-computation.[8]

References

1. Phillippo DM, Dias S, Elsada A, et al. Population Adjustment Methods for Indirect Comparisons: A Review of National Institute for Health and Care Excellence Technology Appraisals. *International journal of technology assessment in health care*. 2019;35(3):221–228. https://doi.org/10.1017/S0266462319000333; 2. Breiman L. Random forests. *Machine learning*. 2001;45:5-32; 3. Rinker TW. wakefield: Generate Random Data. version 0.3.3. 2018. Buffalo, New York. https://github.com/trinker/wakefield; 4. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2022. https://www.R-project.org/; 5. Ishwaran H and Kogalur UB. Fast Unified Random Forests for Survival, Regression, and Classification (RF- SRC), R package version 3.1.1. 2022; Ishwaran H and Kogalur UB. Random survival forests for R. *R News*. 2007;7(2):25-31; Ishwaran H, Kogalur UB, Blackstone EH, and Lauer MS. Random survival forests. *Ann Appl Statist*. 2008;2(3):841-860; 6. Young R. maic: Matching-Adjusted Indirect Comparison_. R package version 0.1.4. 2022. https://CRAN.R-project.org/package=maic; 7. Moradian H, Tremblay G, Senchaudhuri P, Heeg B, et al. MSR11 Evaluating the Impact of Exact Logistic Regression on Matching Adjusted Indirect Comparisons (MAICs) of Treatments Between Trials: A Simulation Study. *Value in Health*. 2022;25(12):S351-S352; 8. Remiro-Azócar A, Heath A, Baio G. Parametric G-computation for compatible indirect treatment comparisons with limited individual patient data. Research Synthesis Methods. 2022;13(6):716-744.

Acknowledgements

The authors would like to thank Stephanie Oldham, Lindsay Weigle, and Colleen Dumont for their writing and editorial support.