Can Low Effective Sample Size in Matching-adjusted Indirect Comparisons Lead to Bias? Findings from a Simulation Study

MSR65

K. Jack Ishak,[1] Conor Chandler,[2] Fei Fei Liu,[3] Sven Klijn[3]

[1]Evidera, Inc., a business unit of PPD, part of Thermo Fisher Scientific, St-Laurent, QC, Canada; [2]Evidera, Inc., a business unit of PPD, part of Thermo Fisher Scientific, Waltham, MA, USA; [3]Bristol-Myers Squibb, Lawrence Township, NJ, USA

Introduction

- Matching-adjusted indirect comparison (MAIC) controls for population differences across trials by weighting patients in the index trial to match the average baseline characteristics of the comparator trial.
- The weighting step in MAIC leads to a loss of precision, which is typically quantified in terms of the effective sample size (ESS).
  - An ESS close to the original sample size indicates that a similar degree of precision is retained; conversely, if the ESS becomes too small, then the precision may be degraded to an unacceptable level.
- The impact of low ESS on the accuracy (i.e., unbiasedness) of estimated treatment effects in MAICs has not been extensively studied, however, and conflicting findings have been published in recent simulation studies for anchored comparisons.[1-3]
  - Phillippo and colleagues found that MAIC may be biased in cases of small sample sizes and large reductions in the ESS.[2]
  - Conversely, a comprehensive simulation study by Remiro-Azocar and colleagues concluded that MAIC yields unbiased treatment effect estimates (under no failure in assumptions), even in analyses that depend on small ESS.[1]
- Therefore, we conducted a simulation study to complement previously published findings and provide practical guidance for the unanchored setting, which is more commonly required in current practice.[4]

Objective

- The primary objective of this simulation study was to determine whether low ESS in MAICs can result in systematic distortions (i.e., bias) in comparative effect estimates, rather than just a loss of precision, despite complete adjustment for imbalanced population characteristics.

Methods

Data-generating mechanisms

- Data were simulated for two single-arm trials with outcomes for an index and comparator treatment to be contrasted with MAIC.
  - The simulated data structure for each population consisted of allocated treatment, one baseline characteristic (continuous covariate), and two outcomes (time to event and dichotomous).
- By design, the true marginal log hazard ratio and log odds ratio for the treatment vs. comparator was 0.

Table 1. Data-generating mechanisms for baseline characteristics and outcomes

Data	Summary of data-generating process
Baseline characteristic (X)	One continuous variable, a PF and EM • Index trial: $X \sim Norm(\mu_1 = 0, \sigma^2 = 1)$ • Comparator trial: $X \sim Norm(\mu_2, \sigma^2 = 1)$
Survival outcome	• Latent event times were simulated for each patient according to a Weibull distribution. • Censoring times were exponentially distributed times based on a common rate parameter assumed to be half of the baseline event rate for all populations and scenarios. • Maximum follow-up for each study was 5 years.
Binary outcome	• Occurrence of an event was defined according to the event status from the simulated survival model, ignoring the time component.

Abbreviations: EM = effect modifier; PF = prognostic factor

Scenarios

- Three parameters were varied in the simulation according to a full factorial design of 3x2x3=18 scenarios.
- The value of each simulation parameter that was tested is presented in **Table 2**, where scenarios were defined to explore all possible interactions.

Table 2. Simulation parameters

Simulation parameters	Levels	Description
Baseline event rate[a]	3	• Low event rate ($\beta_0 = 0.29/5$) • Medium event rate ($\beta_0 = 0.29$) • High event rate ($\beta_0 = 0.29 * 10$)
Population overlap	2	• Poor population overlap: μ_1=0, μ_2=−1.25 • Moderate population overlap: μ_1=0, μ_2=−0.75
Sample size in index trial[b]	3	• Small (n=50 per treatment arm) • Medium (n=125 per treatment arm) • Large (n=250 per treatment arm)

[a]The baseline event rate/scale ($\beta_0 = 0.29$) and shape ($\gamma = 1.27$) were derived by fitting the Weibull distribution to two target quantiles for a hypothetical reference arm: 75% survival at 1 year and 50% survival at 2 years. [b]The sample size in the comparator trial was fixed at n=125.

MAIC

- MAIC balancing weights for the i^{th} subject in the index trial (w_i) represent the conditional log odds that the subject belongs to the comparator population.

$$\ln(w_i) = \ln\left(\frac{\mathbb{P}(\text{Comparator}|x_i)}{1-\mathbb{P}(\text{Comparator}|x_i)}\right) = \alpha + x_i'\beta$$

- The ESS is defined as the original sample size (N) divided by the design effect (D_{eff}), commonly estimated simply as a function of the MAIC balancing weights.

$$\widehat{ESS} = \frac{N}{\widehat{\mathcal{D}}_{eff}} \approx \frac{(\sum_{i=1}^{n}\widehat{w}_i)^2}{\sum_{i=1}^{n}\widehat{w}_i^2}$$

- To estimate the marginal log hazard ratio for treatment, Cox proportional hazards models incorporating w_i were fitted.
- Logistic regression models incorporating w_i were fitted using two methods to estimate the marginal log odds ratio for treatment:
  1. Conventional maximum likelihood estimation, which can be susceptible to small-sample bias
  2. Bias-reduction method (maximum penalized likelihood) developed by Firth.[5]
- Scenarios with a bias of 10% or more were examined to understand the determining factors.

Results

- Bias was typically observed in MAICs of survival and binary outcomes when the absolute ESS was low (<~30 to 35), regardless of whether low ESS resulted from low sample size or poor overlap between populations (**Figures 1 and 2**).
  - In survival outcomes, bias occurred when the effective event count (EEC)—the number of events in the weighted index sample—became very low (<~10).
  - Bias was avoided when the event rate was high, even with low ESS (**Figure 1**).
- For binary outcomes, conventional logistic regression was susceptible to large bias when ESS <30 to 35, particularly when the event rate was low or high (**Figure 2**).
  - Use of a penalized likelihood approach largely addressed this issue, except where the event rate was high (**Figure 3**).

Figure 1. MAICs of survival outcomes based on Cox regression

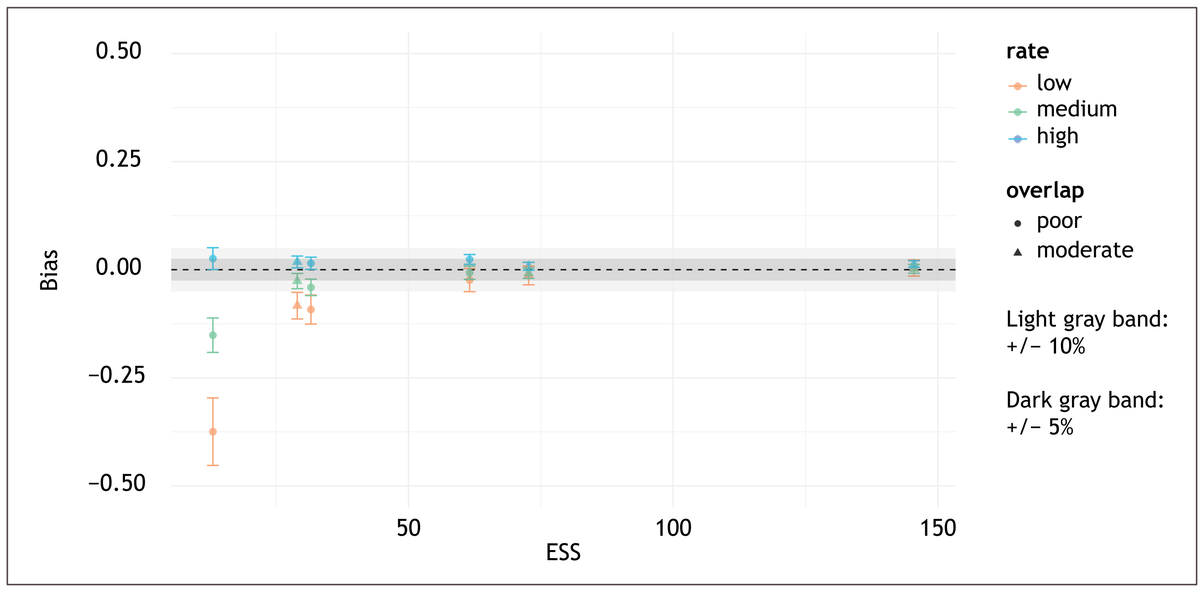


Figure 2. MAICs of binary outcomes based on conventional logistic regression

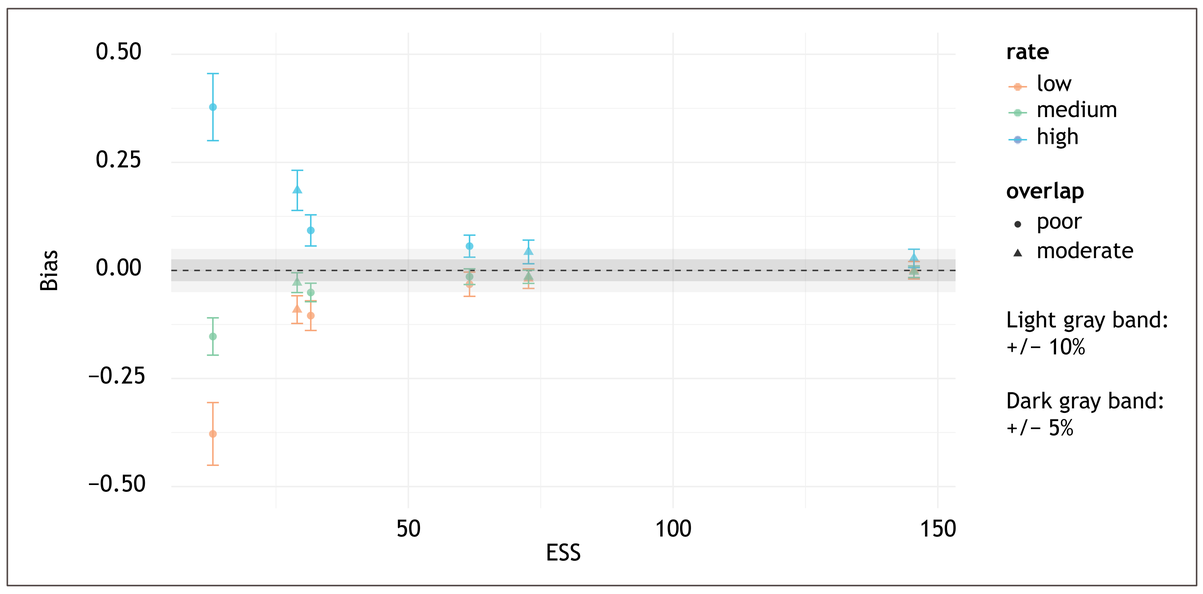
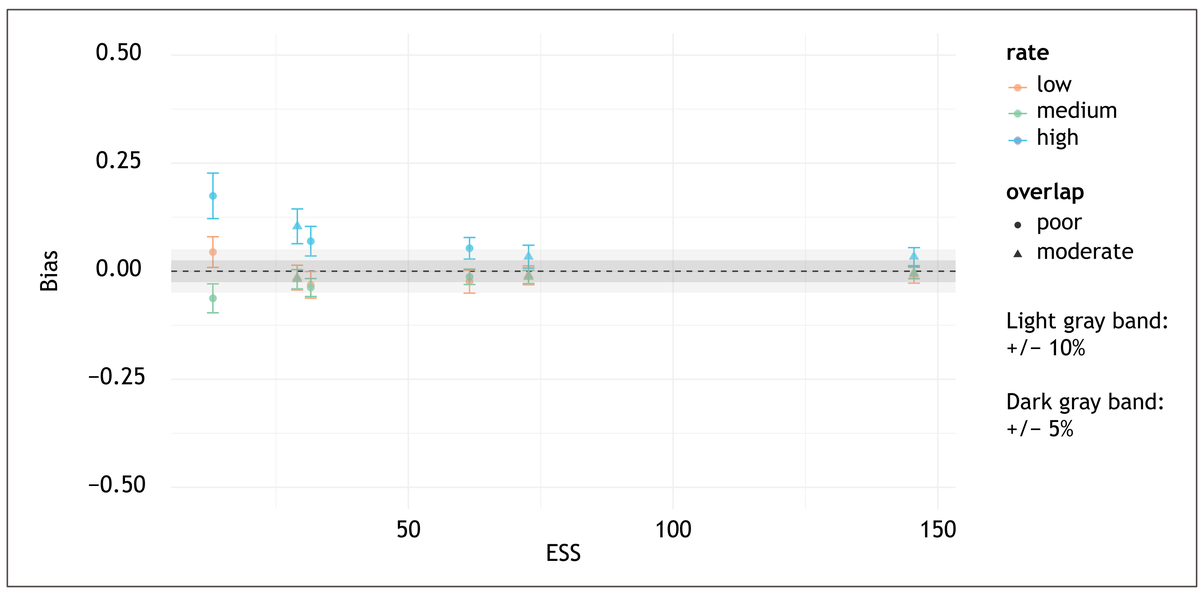


Figure 3. MAICs of binary outcomes based on penalized logistic regression



Conclusions

- Our simulation study confirms the hypothesis that ESS is a useful metric to gauge the potential for bias in MAICs. Caution is advisable when interpreting MAICs where ESS is <~30.
- Low ESS was a necessary but not sufficient condition for bias to occur, however, in the scenarios we investigated.
  - The risk of bias also depends on the type of outcome, statistical methods for fitting the outcome regression model for dichotomous outcomes (conventional vs. penalized likelihood estimation), and the underlying event rate.
- In practice, analysts should first consider the type of outcome and underlying event rate:
  - For survival outcomes, MAICs were only prone to bias in situations where the event rate was not high and matching led to a low ESS. When the event rate is high, the number of events contributing to the analysis after matching may remain relatively high and prevent distortions.
  - For binary outcomes, analyses of high event rates (i.e., ≥75%) with a low ESS should also be cause for concern, as the bias was estimated to be similar in magnitude to scenarios of low event rates. That is, in contrast to MAICs for survival outcomes, low ESS is always problematic in MAICs of binary outcomes, regardless of how it occurred.
- MAICs typically compare outcomes derived using conventional maximum likelihood estimation, which can be susceptible to small-sample bias.
  - MAICs of binary outcomes based on bias-reduced logistic regression resulted in improved accuracy and precision compared with conventional methods, particularly in low-rate scenarios.
  - Using a bias-reduction method, such as Firth's penalized likelihood for logistic regression or Cox proportional hazards models, is recommended.

References

1. Remiro-Azocar A, et al. *Res Synth Methods*. 2021;12(6):750-775.
2. Phillippo DM, et al. *Stat Med*. 2020;39(30):4885-4911.
3. Remiro-Azocar A, et al. *Res Synth Methods*. 2022;13(6):716-744.
4. Phillippo DM, et al. *Int J Technol Assess Health Care*. 2019;35(3):221-228.
5. Firth D. *Biometrika*. 1993;80:27-38.

Acknowledgments

• Editorial and graphic design support were provided by Michael Grossi, Kawthar Nakayima, and Shani Berger of Evidera Inc., a business unit of PPD, part of Thermo Fisher Scientific.

Presented at ISPOR Europe 2024; November 17-20, 2024; Barcelona, Spain

Email: Jack.Ishak@evidera.com