Uncertain About Uncertainty in Matching-Adjusted Indirect Comparisons (MAICs)? A Simulation Study to Compare Methods for Variance Estimation

Conor Chandler[1], Irina Proskorovsky[2]

[1]Evidera, Waltham, MA, USA; [2]Evidera, Spain

Introduction

- Indirect treatment comparisons (ITCs) are commonly required in health technology assessments (HTAs) to compare the efficacy and safety of all competing interventions that have not already been studied head-to-head in a clinical trial
  - Treatment effect estimates from ITCs are typically key drivers of results in cost-effectiveness analyses, providing a substantial contribution to the evidence base for reimbursement recommendations
  - Accurately estimating the uncertainty around treatment effects is critical to ensuring robust data are available to guide healthcare decision-making
- In cases where there are noteworthy imbalances in baseline characteristics across trials considered in ITCs, HTA guidelines (e.g., NICE,[1] EUnetHTA[2]) recommend population-adjusted indirect comparisons (PAICs)
  - Matching-adjusted indirect comparison (MAIC) is the most common PAIC methodology considered in technology appraisals
- The MAIC method relies on propensity score weighting to adjust for confounding
  - Previous studies[3,4] have demonstrated that conventional methods for variance estimation are biased in propensity score-weighted analyses, typically understating the true uncertainty, and alternative methods have been proposed to correct for this downward bias
- The NICE decision support unit (DSU) recommends using the sandwich estimator or bootstrapping to derive the variance of estimates from MAICs[1]
- To date, no simulation studies have been conducted to investigate and compare the performance of different methods to estimate uncertainty in the context of MAIC
  - Because different methods for variance estimation may lead to a range of cost-effectiveness results and possible inconsistencies in HTAs, additional guidance and recommendations are required

Objective

- The aim of this simulation study was to assess the performance of different statistical methods for estimating the uncertainty around treatment effects derived via an anchored MAIC

Methods

- Monte Carlo simulations (N=1,000 replications) were conducted for an anchored MAIC of two trials using the original method proposed by Signorovitch et al., 2010[5] and recommended by the NICE DSU[1]

Data generation for baseline characteristics

- Assumed an imbalance in two effect modifiers (EMs) (that were also prognostic) between trials
- For the index trial, baseline characteristics were simulated according to the same distributions across all scenarios, whereas distributions for the comparator trial were varied to induce different levels of population overlap (Table 1)

Table 1. Baseline Characteristics

Baseline Characteristics	Index Trial ($k = 1$)	Comparator Trial ($k = 2$)
X_1 (continuous variable)	$X_1 \sim Norm(\mu_{k=1} = 10, \sigma^2 = 2^2)$	$X_1 \sim Norm(\mu_{k=2}, \sigma^2 = 2^2)$
X_2 (binary variable)	$X_2 \sim Binom(p_{k=1} = 0.5)$	$X_2 \sim Binom(p_{k=2})$

Baseline characteristics were simulated from a multivariate Gaussian copula with moderate positive correlation ($r=0.5$) in each trial

Scenarios

- A total of 18 scenarios were conducted to investigate the impact of outcome type, sample size in the index trial, and population overlap on variance estimation in MAICs (Table 2)

Table 2. Simulation Scenarios

Parameter(s) to Vary	Number of Levels	Description
Type of outcome	2	• Survival • Binary
Total sample size in index trial	3	• $N \in \{200, 400, 800\}$
Population overlap	3	• $\mu_{k=2} \in \{12.5, 11.5, 10.5\}$ and $p_{k=2} \in \{0.7, 0.6, 0.55\}$ (poor, moderate, strong)

The total sample size for the comparator trial was fixed at N=200 (100 patients per treatment arm)

Data generation for outcomes

- Assumed that the binary response variable measured a positive event (e.g., response), whereas survival was modeled as a negative event (e.g., overall survival)
  - Active treatments had favorable relative effects vs. control arms in both index and comparator trials
- Regression parameters were calibrated to achieve target milestones, including the following:
  - Binary and survival events experienced by 30% and 50% of the control arm and by 50% and 40% of the active arm in the index population
  - Median survival of 60 and 100 months for the control and active arms, respectively, in the index population

Statistical methods

- Weighted logistic and Cox proportional hazards models were fitted to estimate log odds ratios and log hazard ratios
- Four different methods were used to estimate the model-based variance:
  1. Conventional estimators (CEs) using raw weights
  2. CEs using weights rescaled to the effective sample size (ESS)
  3. Robust sandwich estimators (HC3 estimator)
  4. Ordinary nonparametric bootstrap with replacement (1,000 resamples)
- The performance of each method was evaluated on the basis of empirical coverage of 95% confidence intervals (CIs) and the ratio of average estimated standard error (SE) versus empirical SE

Results

- The empirical coverage probabilities for CE + ESS-scaled weights (ranging from 94.5% to 96.4%) did not deviate significantly from the nominal confidence level in 17 of 18 scenarios (Figure 1)
- On the contrary, variance was underestimated by CE + raw weights (6/6 scenarios), bootstrapping (4/6), and sandwich estimators (3/6) in the scenarios with poor population overlap (Figure 2)
- The use of CE + raw weights underestimated variance in half of the scenarios with moderate population overlap, whereas all other methods had unbiased results
- All four methods provided accurate estimates of the variance in scenarios with strong population overlap, with empirical coverage probabilities ranging from 0.97 to 1.06 and SE ratios between 94.2% and 96.2%

Figure 1. Comparison of Empirical Coverage Probabilities Across Scenarios
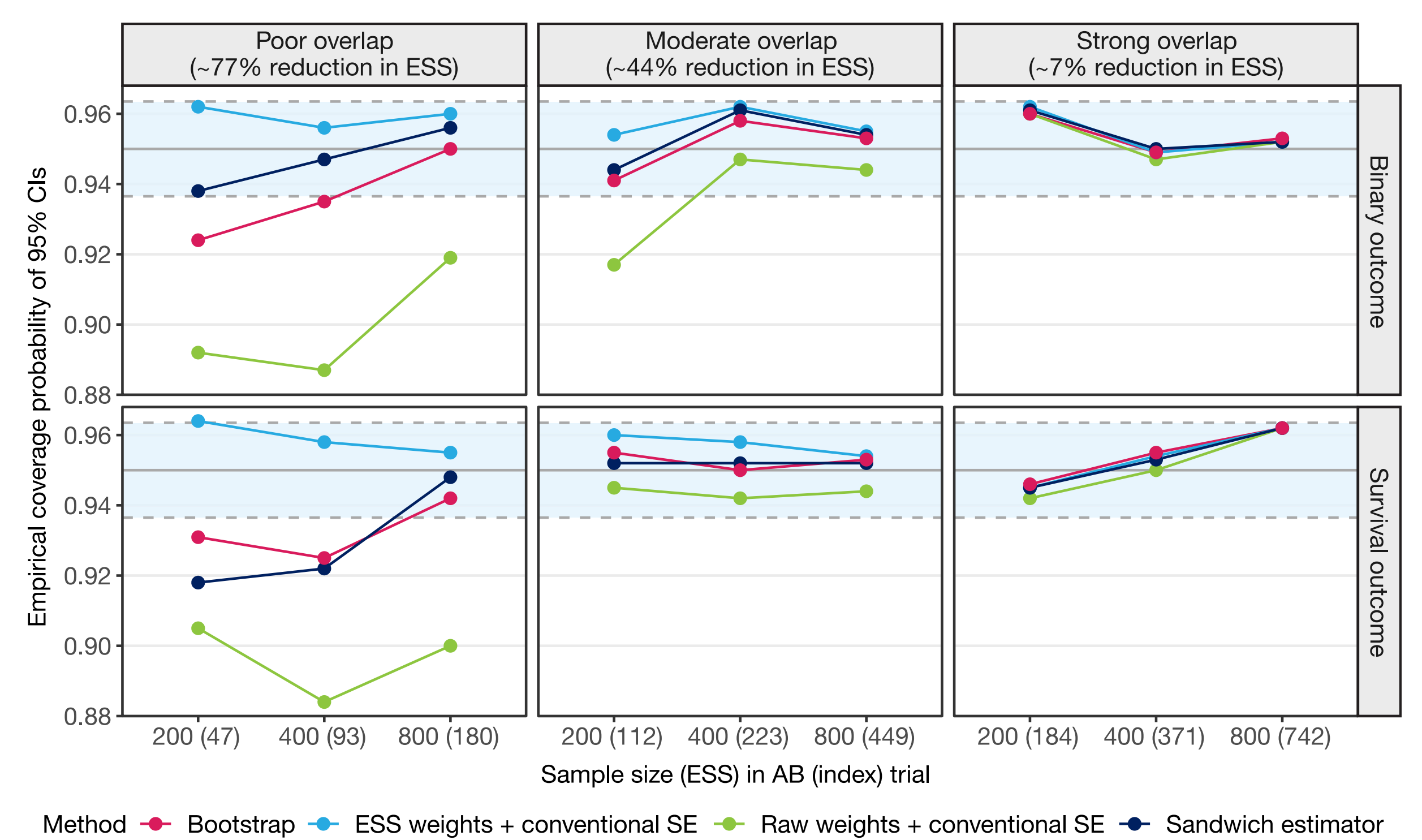
The shaded band represents the 95% CI. Estimates within this shaded region did not significantly differ from the nominal level of 95%. For bootstrapping, the SE was estimated based on 1,000 resamples, and asymptomatic CIs were constructed assuming a normal distribution. Abbreviations: CI = confidence interval; ESS = effective sample size; SE = standard error

Figure 2. Comparison of Variability Ratios Across Scenarios
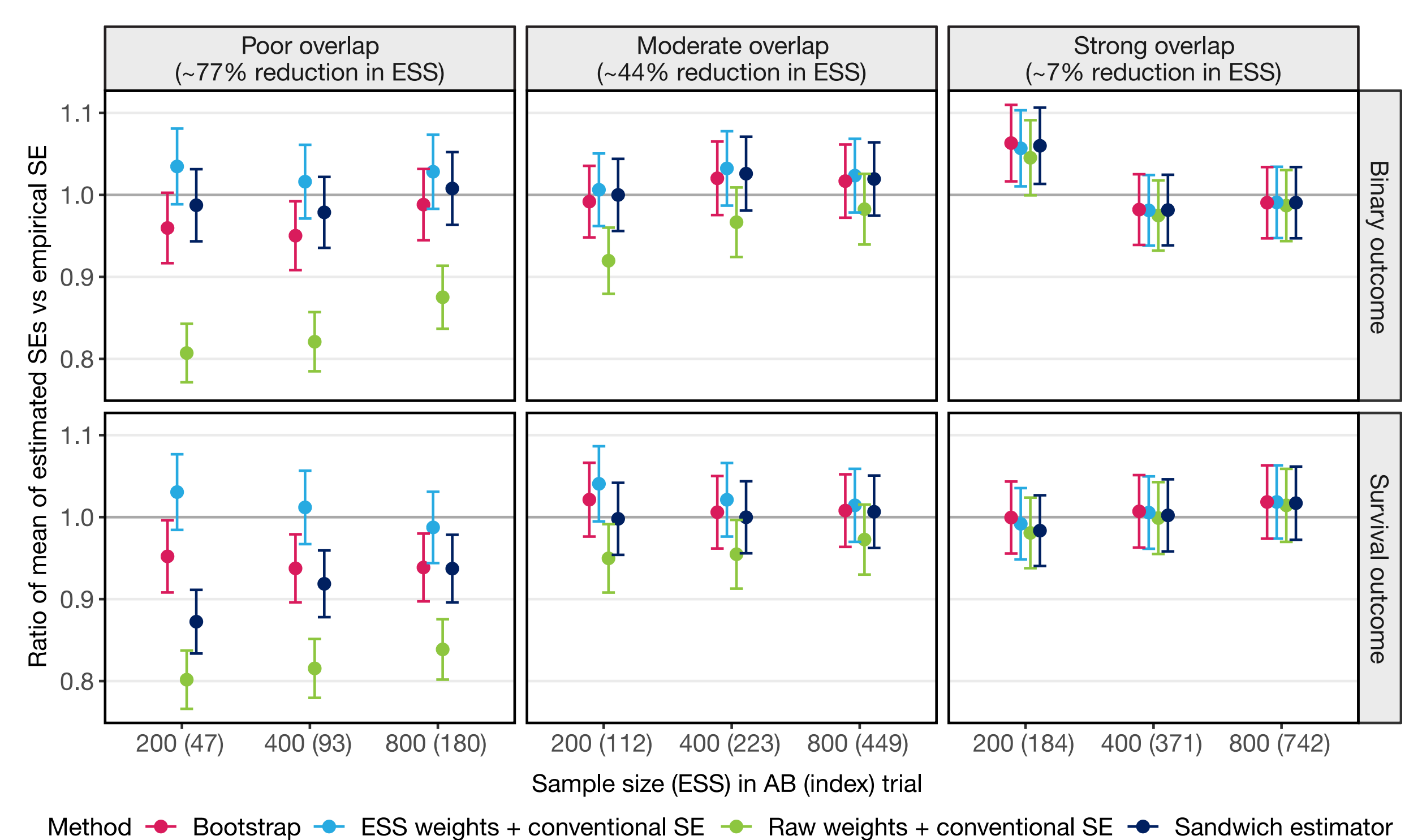
Abbreviations: ESS = effective sample size; SE = standard error

Discussion

- Current guidance recommends against the use of conventional variance estimators in MAICs, but this simulation study demonstrates that conventional estimators can be accurate if the weights are correctly scaled to reflect the ESS
- The idealized scenarios presented herein assumed all EMs were accounted for, in which case MAICs would be theoretically unbiased under no failure in assumptions; in practice, however, any residual bias would inevitably alter the performance of these methods (e.g., reduce coverage probability)
- These simulations considered only one version of the robust sandwich estimator
  - There are a variety of robust sandwich estimators (e.g., HC1, HC2, HC3), and the defaults may differ by statistical software
- Ordinary bootstrapping underestimated the variance in scenarios with poor overlap, likely due to small sample bias
  - Alternative bootstrapping methods (e.g., bias-correction) may provide more accurate estimates of CIs
- The generalizability of these findings to unanchored MAICs is unclear and requires additional research

Conclusions

- The extent of population overlap is an important consideration for variance estimation in MAICs
- Variance estimates based on CE + raw weights were systematically biased downwards; thus, this method should be avoided
- Use of CE + ESS-scaled weights produced SEs and CIs that were precise across all scenarios, whereas robust sandwich estimators and bootstrapping underperformed in several scenarios with poor overlap
- Future research is needed to supplement findings and continue to investigate the performance of variance estimation methods under alternative scenarios (e.g., unanchored MAIC)

References

1. Phillippo D, et al. NICE DSU technical support document 18. December 2016.   2. EUnetHTA. Methods guideline D4.3.2. August 2022.   3. Austin PC. *Stat Med*. 2016;35(30):5642-5655.   4. Robins JM, et al. *Epidemiology*. 2000;11(5):550-560.   5. Signorovitch JE, et al. *Pharmacoeconomics*. 2010;28(10):935-945.