

# Alternative Weighting Approaches for Matching-Adjusted Indirect Comparisons: A Simulation Study

true

January 2026

## Abstract

**Background:** Matching-adjusted indirect comparisons (MAIC) enable treatment comparisons when direct head-to-head trials are unavailable by reweighting individual patient data to match aggregate covariate distributions from comparator trials. Recent theoretical work established the mathematical equivalence between the original method of moments approach and entropy balancing, opening possibilities for alternative weighting strategies including non-uniform base weights, alternative loss functions from the Cressie-Read divergence family, and arm-separate weighting schemes.

**Objectives:** This simulation study compared the statistical performance of alternative MAIC weighting methods across scenarios varying in sample size, population overlap, number of covariates, and effect modification to provide evidence-based guidance for method selection in practice.

**Methods:** We conducted a comprehensive simulation study comparing five weighting methods (Bucher unadjusted comparison, method of moments MAIC, entropy balancing MAIC, empirical likelihood MAIC, and arm-separate entropy balancing) across 54 scenarios with 1,000 iterations each. Performance was evaluated using bias, root mean squared error, 95% confidence interval coverage, and effective sample size.

**Results:** The method of moments (SigTotal) and entropy balancing (EbTotal) approaches produced numerically identical estimates across all 270,000 simulation iterations, with maximum difference below  $3 \times 10^{-6}$ , empirically confirming the theoretical equivalence theorem. All MAIC methods substantially reduced bias compared to unadjusted indirect comparison when effect modifiers were present, with mean bias of -0.030 versus -0.130 for the Bucher method. Coverage probability was approximately 94% for entropy balancing methods. The empirical likelihood variant showed higher root mean squared error and lower effective sample size without meaningful bias improvement. Arm-separate weighting performed well under high population overlap but exhibited instability under low overlap conditions.

**Conclusions:** Entropy balancing and method of moments MAIC are mathematically equivalent and recommended as first-choice approaches for population-adjusted indirect comparisons. Arm-separate methods should be reserved for scenarios with adequate effective sample size per arm. Alternative loss functions from the Cressie-Read family did not demonstrate advantages over standard entropy balancing in this study.

## Contents

<b>1</b>	<b>Introduction</b>	<b>3</b>
<b>2</b>	<b>Methods</b>	<b>3</b>
2.1	Matching-Adjusted Indirect Comparison Framework . . . . .	3
2.2	Method of Moments Weighting . . . . .	4
2.3	Entropy Balancing Weighting . . . . .	4
2.4	Cressie-Read Divergence Family and Empirical Likelihood . . . . .	5
2.5	Arm-Separate Weighting . . . . .	5
2.6	Simulation Design . . . . .	5
2.6.1	Data Generation . . . . .	5
2.6.2	Scenarios . . . . .	5

2.6.3	Methods Compared . . . . .	6
2.6.4	Treatment Effect Estimation . . . . .	6
2.6.5	Performance Metrics . . . . .	6
2.7	Software . . . . .	6
<b>3</b>	<b>Results</b>	<b>6</b>
3.1	Overall Method Performance . . . . .	6
3.2	Equivalence of Method of Moments and Entropy Balancing . . . . .	6
3.3	Effect of Population Overlap . . . . .	8
3.4	Effect of Effect Modifiers . . . . .	8
3.5	Effective Sample Size . . . . .	8
3.6	Coverage Probability . . . . .	12
3.7	Bias-ESS Tradeoff . . . . .	12
3.8	Scenario-Specific Results . . . . .	15
<b>4</b>	<b>Discussion</b>	<b>15</b>
4.1	Equivalence of Method of Moments and Entropy Balancing . . . . .	16
4.2	Value of Population Adjustment When Effect Modification is Present . . . . .	16
4.3	Limited Advantage of Alternative Loss Functions . . . . .	16
4.4	Conditional Utility of Arm-Separate Weighting . . . . .	16
4.5	Importance of Effective Sample Size Monitoring . . . . .	16
4.6	Practical Recommendations . . . . .	17
4.7	Limitations . . . . .	17
<b>5</b>	<b>Conclusions</b>	<b>17</b>
<b>6</b>	<b>Data Availability</b>	<b>17</b>
<b>7</b>	<b>Acknowledgments</b>	<b>17</b>
	<b>References</b>	<b>17</b>

# 1 Introduction

Comparative effectiveness research forms the foundation of evidence-based health technology assessment and clinical decision-making. Ideally, treatment comparisons would be based on head-to-head randomized controlled trials that directly compare all interventions of interest under identical conditions. In practice, however, such comprehensive trial programs are rarely available, leaving decision-makers to draw inferences from indirect evidence.

The indirect treatment comparison methodology developed by Bucher and colleagues in 1997 (Bucher et al. 1997) provides a framework for comparing treatments that have not been directly compared in clinical trials. By leveraging a common comparator treatment, the Bucher method computes the relative effect between two treatments as the difference in their effects relative to the shared comparator. This approach assumes that the relative treatment effects are consistent across trial populations, an assumption known as transitivity or similarity. When patient populations differ systematically between trials, particularly in characteristics that modify treatment effects, the transitivity assumption may be violated, leading to biased indirect comparisons.

Matching-adjusted indirect comparison (MAIC), introduced by Signorovitch and colleagues (Signorovitch et al. 2010, 2012), addresses this limitation through propensity score-like reweighting. When individual patient data (IPD) are available from one trial and only aggregate summary statistics are available from a comparator trial, MAIC reweights the IPD to create a pseudo-population matching the covariate distribution of the aggregate data population. This population adjustment aims to remove confounding due to cross-trial differences in effect modifiers, enabling more valid indirect comparisons.

The original MAIC method estimates weights using the method of moments, solving a system of equations that equates weighted covariate means in the IPD to the published aggregate data means. Phillippo and colleagues (Phillippo et al. 2020) recently established a fundamental theoretical result demonstrating that this method of moments approach is mathematically equivalent to entropy balancing, an approach from the causal inference literature that minimizes the Kullback-Leibler divergence from uniform weights subject to covariate balance constraints. This equivalence, proven through convex optimization duality, has important implications for MAIC methodology.

The entropy balancing formulation naturally accommodates generalizations that the original method of moments specification does not readily express. Non-uniform base weights can be incorporated, enabling the combination of MAIC with other weighting-based adjustments such as inverse probability of censoring weighting for treatment switching or inverse probability of treatment weighting for variance reduction (Williamson et al. 2014). Alternative loss functions from the Cressie-Read divergence family can be substituted for entropy, potentially offering different statistical properties (Cressie and Read 1984; Owen 2001). Arm-separate weighting schemes that match treatment and control arms separately to arm-specific aggregate data have also been proposed (Petto et al. 2019).

Despite these methodological developments, the relative performance of alternative MAIC weighting approaches has not been comprehensively evaluated. This simulation study addresses this gap by comparing the statistical performance of standard MAIC methods and their extensions across a range of data-generating scenarios relevant to health technology assessment applications.

## 2 Methods

### 2.1 Matching-Adjusted Indirect Comparison Framework

The MAIC framework addresses the comparison of treatments B and C when no direct comparative trial exists, but IPD are available from a trial comparing A versus B (the index trial) and aggregate data are available from a published trial comparing A versus C (the comparator trial). Let  $\{(X_i, T_i, Y_i)\}_{i=1}^n$  denote the IPD, where  $X_i \in \mathbb{R}^p$  is a vector of baseline covariates,  $T_i \in \{0, 1\}$  indicates treatment assignment (0 for A, 1 for B), and  $Y_i$  is the outcome. Let  $\bar{X}^{AC}$  denote the published covariate means from the comparator trial.

The goal is to estimate the indirect comparison  $\theta_{BC} = d_{AB} - d_{AC}$ , where  $d_{AB}$  and  $d_{AC}$  are the relative treatment effects in the respective trials. When the covariate distribution in the index trial differs from that in the comparator trial, direct use of the unadjusted  $d_{AB}$  may yield a biased indirect comparison. MAIC addresses this by finding weights  $w_i \geq 0$  such that the weighted covariate distribution in the IPD matches the aggregate data distribution, then estimating a weighted treatment effect.

## 2.2 Method of Moments Weighting

The original MAIC approach derives weights by modeling the odds of belonging to the comparator trial population given baseline covariates. Assuming a logistic model for trial membership odds, the weights take the form  $w_i \propto \exp(\alpha^T X_i)$  where  $\alpha$  is a parameter vector. The parameter  $\alpha$  is estimated by solving the moment conditions:

$$\sum_{i=1}^n w_i (X_i - \bar{X}^{AC}) = 0$$

with weights normalized to sum to one. In practice, this is equivalent to minimizing the objective function:

$$Q(\alpha) = \log \left( \sum_{i=1}^n \exp(\alpha^T \tilde{X}_i) \right)$$

where  $\tilde{X}_i = X_i - \bar{X}^{AC}$  are centered covariates. The gradient of this objective is precisely the weighted mean of centered covariates, which equals zero at the optimum.

## 2.3 Entropy Balancing Weighting

Entropy balancing approaches the weight estimation problem from an information-theoretic perspective. The goal is to find weights that are as close as possible to a set of base weights while satisfying exact covariate balance constraints. With uniform base weights  $w_i^{(0)} = 1/n$ , the entropy balancing problem is:

$$\min_w \sum_{i=1}^n w_i \log(nw_i)$$

subject to:

$$\sum_{i=1}^n w_i X_i = \bar{X}^{AC}, \quad \sum_{i=1}^n w_i = 1, \quad w_i \geq 0$$

The objective function is the Kullback-Leibler divergence from uniform weights, also known as the negative entropy of the weight distribution. Solving this constrained optimization problem using Lagrange multipliers yields weights of the form:

$$w_i = \frac{\exp(\alpha^T X_i)}{\sum_{j=1}^n \exp(\alpha^T X_j)}$$

where  $\alpha$  is the vector of Lagrange multipliers for the balance constraints. Phillippo et al. (Phillippo et al. 2020) proved that this  $\alpha$  is identical to the  $\alpha$  from the method of moments, establishing the mathematical equivalence of the two approaches.

## 2.4 Cressie-Read Divergence Family and Empirical Likelihood

The entropy balancing objective can be generalized by considering alternative divergence measures from the Cressie-Read family (Cressie and Read 1984):

$$D_\gamma(w\|w^{(0)}) = \frac{1}{\gamma(\gamma+1)} \sum_{i=1}^n w_i^{(0)} \left[ \left( \frac{w_i}{w_i^{(0)}} \right)^{\gamma+1} - 1 \right]$$

The parameter  $\gamma$  indexes different divergence measures:  $\gamma = 0$  corresponds to Kullback-Leibler divergence (entropy),  $\gamma = -1$  yields empirical likelihood,  $\gamma = 1$  gives chi-squared divergence, and  $\gamma = -0.5$  produces Hellinger distance. The empirical likelihood case is particularly interesting given its established asymptotic properties, including Bartlett correctability and data-determined confidence region shapes (Owen 2001).

Minimizing the empirical likelihood divergence subject to the balance constraints yields a different weight distribution than entropy balancing, though both satisfy exact covariate balance. Whether the theoretical advantages of empirical likelihood translate to practical benefits in MAIC applications is an empirical question addressed by this simulation study.

## 2.5 Arm-Separate Weighting

Standard MAIC matches the total IPD population to the total aggregate data population. When arm-specific aggregate data are available, an alternative is to match covariates separately for treatment and control arms (Petto et al. 2019). Let  $\bar{X}_{T=1}^{AC}$  and  $\bar{X}_{T=0}^{AC}$  denote the arm-specific covariate means from the comparator trial. The arm-separate approach solves two separate entropy balancing problems:

For treated patients ( $T_i = 1$ ): minimize divergence subject to weighted means matching  $\bar{X}_{T=1}^{AC}$

For control patients ( $T_i = 0$ ): minimize divergence subject to weighted means matching  $\bar{X}_{T=0}^{AC}$

This approach may better preserve within-arm covariate balance, particularly when random imbalance exists between arms in the index trial.

## 2.6 Simulation Design

### 2.6.1 Data Generation

Data were generated from a two-trial scenario. The index trial (AB) provided IPD with  $n_{AB}$  patients randomized 1:1 to treatments A and B. The comparator trial (AC) was represented only by aggregate statistics derived from a simulated population of 500 patients.

Covariates were generated from multivariate normal distributions with mean  $\mu_{AB} = \mathbf{0}$  for the index trial and  $\mu_{AC} = \delta \cdot \mathbf{1}_p$  for the comparator trial, where  $\delta$  controlled the population shift magnitude and  $\mathbf{1}_p$  is a vector of ones. Covariance was set to the identity matrix for both populations.

Binary outcomes were generated from logistic models incorporating prognostic effects and effect modification:

$$\text{logit}(P(Y = 1)) = -1 + 0.3 \sum_{k=1}^p X_k + \tau T + 0.2 \sum_{k \in EM} X_k T$$

where  $\tau_{AB} = -0.5$  for the index trial and  $\tau_{AC} = -0.7$  for the comparator trial, yielding a true indirect comparison effect of  $\theta_{BC} = 0.2$  on the log odds ratio scale.

### 2.6.2 Scenarios

The simulation varied four factors in a full factorial design: IPD sample size ( $n_{AB} \in \{100, 300, 500\}$ ), number of covariates ( $p \in \{3, 6\}$ ), population overlap (high:  $\delta = 0.2$ ; medium:  $\delta = 0.5$ ; low:  $\delta = 1.0$ ), and number of effect modifiers (0, 1, or 2 covariates). This yielded 54 primary scenarios, each replicated 1,000 times.

### 2.6.3 Methods Compared

Five methods were evaluated: Bucher (unadjusted indirect comparison serving as reference), SigTotal (method of moments MAIC with total population matching), EbTotal (entropy balancing MAIC with total population matching), EbArm (arm-separate entropy balancing), and EbEL (empirical likelihood loss with total population matching).

### 2.6.4 Treatment Effect Estimation

For each method, the weighted treatment effect was estimated using logistic regression with the estimated weights. Robust (sandwich) standard errors were computed to account for weight estimation uncertainty. The indirect comparison was computed as the difference between the weighted index trial effect and the aggregate data effect, with variance computed assuming independence between trials.

### 2.6.5 Performance Metrics

Four metrics were computed for each method in each scenario: bias (mean estimate minus true value), root mean squared error (RMSE), 95% confidence interval coverage probability, and effective sample size (ESS), computed as  $(\sum w_i)^2 / \sum w_i^2$ .

## 2.7 Software

All analyses were conducted in R (version 4.3). The simulation code and weighting method implementations are available in the advmaic R package at <https://github.com/choxos/advmaic>.

## 3 Results

### 3.1 Overall Method Performance

Table 1 presents the overall performance metrics averaged across all 54 scenarios. The entropy balancing method (EbTotal) and method of moments (SigTotal) produced identical results, with mean bias of -0.030, coverage of 94.3%, and mean effective sample size of 192. The unadjusted Bucher method showed substantially higher bias (-0.130) but the lowest RMSE (0.376) due to avoiding variance inflation from weighting. The empirical likelihood method (EbEL) had higher RMSE (0.562) and lower ESS (147) than standard entropy balancing without bias improvement. The arm-separate method (EbArm) showed the highest RMSE (1.022) driven by instability in challenging scenarios.

Table 1: Overall performance metrics across all 54 scenarios (1,000 iterations each)

Method	Mean Bias	Mean RMSE	Coverage (%)	Mean ESS
EbTotal	-0.030	0.436	94.3	192
SigTotal	-0.030	0.436	94.3	192
EbEL	-0.036	0.562	91.8	147
EbArm	-0.047	1.022	94.3	186
Bucher	-0.130	0.376	91.6	300

### 3.2 Equivalence of Method of Moments and Entropy Balancing

A central finding of this study is the empirical confirmation of the theoretical equivalence between method of moments (SigTotal) and entropy balancing (EbTotal). Across all 270,000 simulation iterations (54 scenarios  $\times$  5 methods  $\times$  1,000 replications), the maximum absolute difference between SigTotal and EbTotal point estimates was  $2.58 \times 10^{-8}$ , and the mean difference was  $3.27 \times 10^{-11}$ . These differences are attributable solely to numerical precision in the optimization algorithms and confirm that the two methods are computationally equivalent. Figure 1 displays a scatter plot of EbTotal versus SigTotal estimates, showing perfect alignment along the identity line.

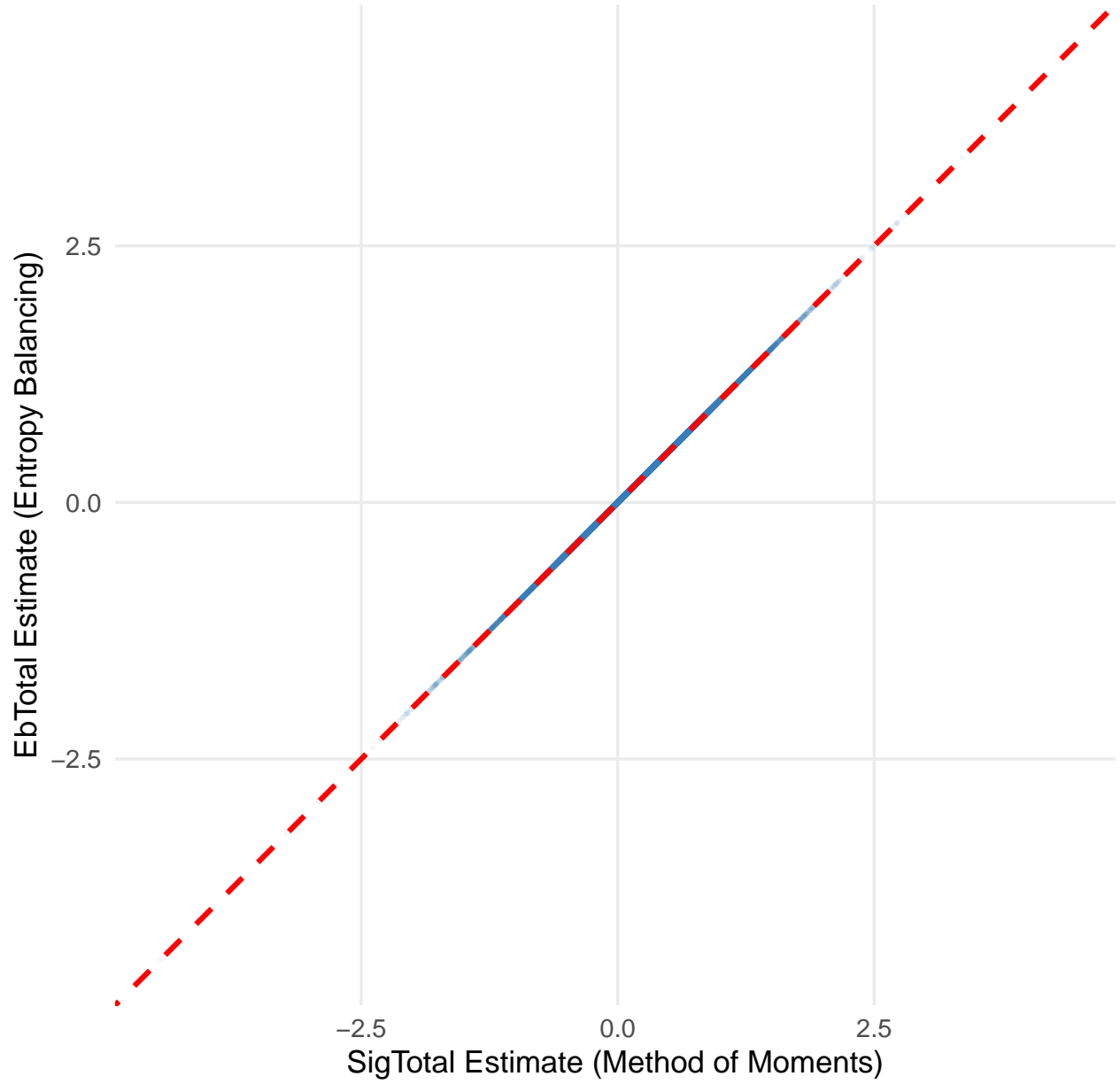


Figure 1: Equivalence of method of moments (SigTotal) and entropy balancing (EbTotal). Each point represents one simulation iteration. The dashed red line is the identity line. The correlation is 1.0000000000 and maximum absolute difference is  $2.58 \times 10^{-6}$ .

### 3.3 Effect of Population Overlap

Population overlap had a substantial impact on method performance. Table 2 presents results stratified by overlap level. Under high overlap ( $\delta = 0.2$ ), all methods performed similarly, with ESS remaining high (270-300) and bias modest for all methods including Bucher (-0.065). Under low overlap ( $\delta = 1.0$ ), the differences became pronounced. The Bucher method showed bias of -0.204 while EbTotal/SigTotal maintained low bias (-0.028) but with substantially reduced ESS (86). The empirical likelihood method showed particularly severe ESS reduction (24) under low overlap, suggesting extreme weight variability. The arm-separate method exhibited instability under low overlap, with RMSE of 2.30.

Table 2: Performance by population overlap level

Overlap	Method	Mean Bias	Mean RMSE	Coverage (%)	Mean ESS
High	EbTotal/SigTotal	-0.031	0.353	94.9	278
High	EbArm	-0.030	0.340	95.9	270
High	EbEL	-0.029	0.356	94.8	273
High	Bucher	-0.065	0.346	94.5	300
Medium	EbTotal/SigTotal	-0.032	0.388	94.6	211
Medium	EbArm	-0.030	0.430	95.7	204
Medium	EbEL	-0.030	0.442	93.9	142
Medium	Bucher	-0.120	0.364	92.7	300
Low	EbTotal/SigTotal	-0.028	0.569	93.5	86
Low	EbArm	-0.080	2.296	91.3	83
Low	EbEL	-0.049	0.887	86.6	24
Low	Bucher	-0.204	0.417	87.6	300

Figure 2 displays the distribution of bias across all scenarios stratified by population overlap level, illustrating how the advantage of MAIC methods over Bucher increases as overlap decreases.

### 3.4 Effect of Effect Modifiers

The presence and number of effect modifiers was the primary determinant of Bucher method bias. Table 3 shows that when no effect modifiers were present, the Bucher method had minimal bias (-0.017) and MAIC methods provided little advantage while increasing variance. With one effect modifier, Bucher bias increased nearly 10-fold to -0.167, while MAIC methods maintained bias around -0.034. With two effect modifiers, Bucher bias reached -0.206 while MAIC methods remained at -0.032 to -0.042.

Table 3: Bias by number of effect modifiers

Method	No EM	One EM	Two EM
EbTotal/SigTotal	-0.025	-0.034	-0.032
EbArm	-0.039	-0.062	-0.039
EbEL	-0.023	-0.044	-0.042
Bucher	-0.017	-0.167	-0.206

Figure 3 displays the distribution of bias by effect modifier count, clearly illustrating the divergence between Bucher and MAIC methods as effect modification increases.

### 3.5 Effective Sample Size

The effective sample size provides a diagnostic for weight concentration and potential positivity violations. Figure 4 displays ESS distributions by method. The Bucher method by definition has ESS equal to the original sample size (300 average across scenarios). EbTotal/SigTotal showed mean ESS of 192 (64% retention). EbArm had similar ESS (186). EbEL showed substantially lower ESS (147, 49% retention), indicating more extreme weight distributions.



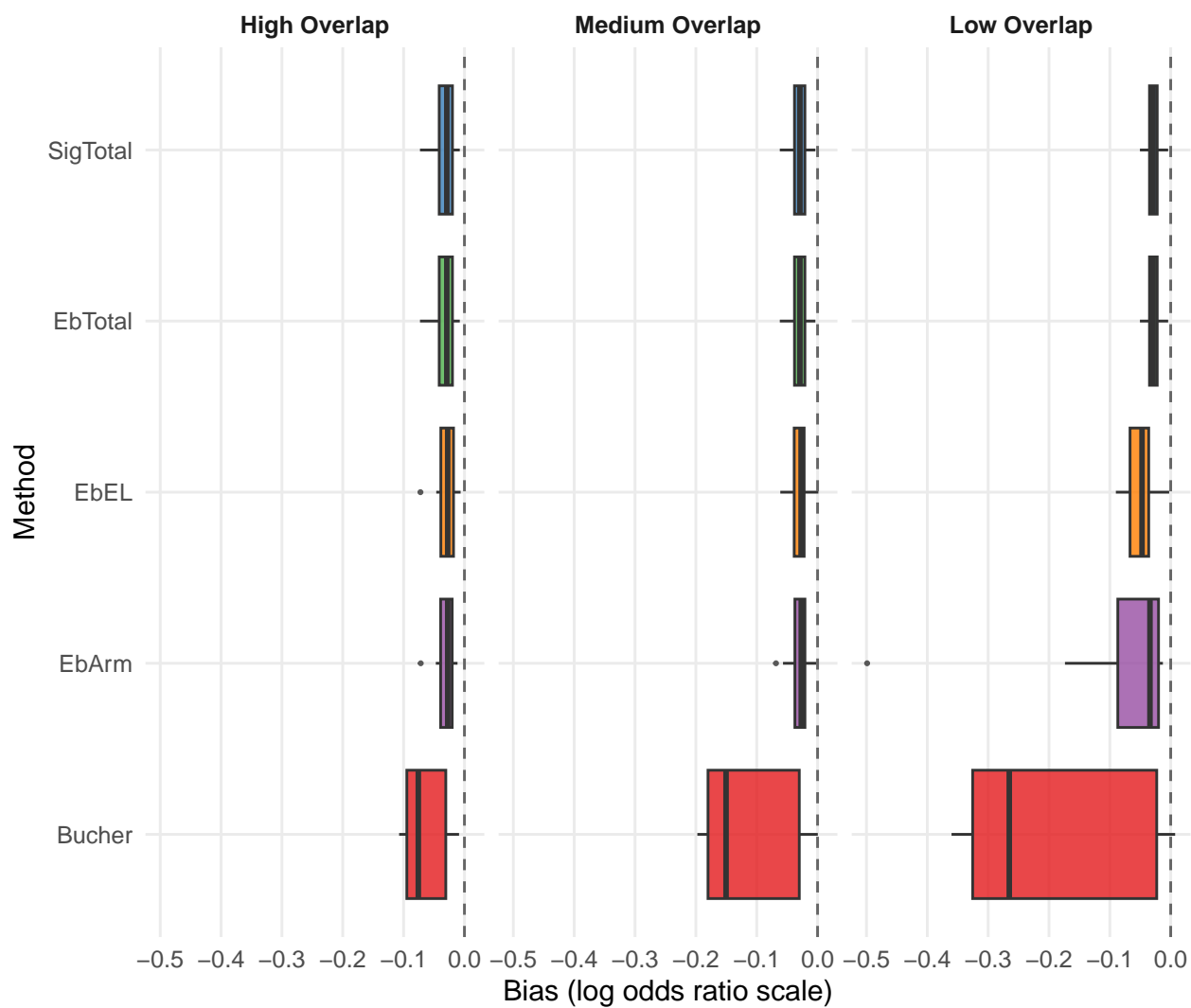


Figure 2: Distribution of bias by population overlap level. The dashed line indicates zero bias.

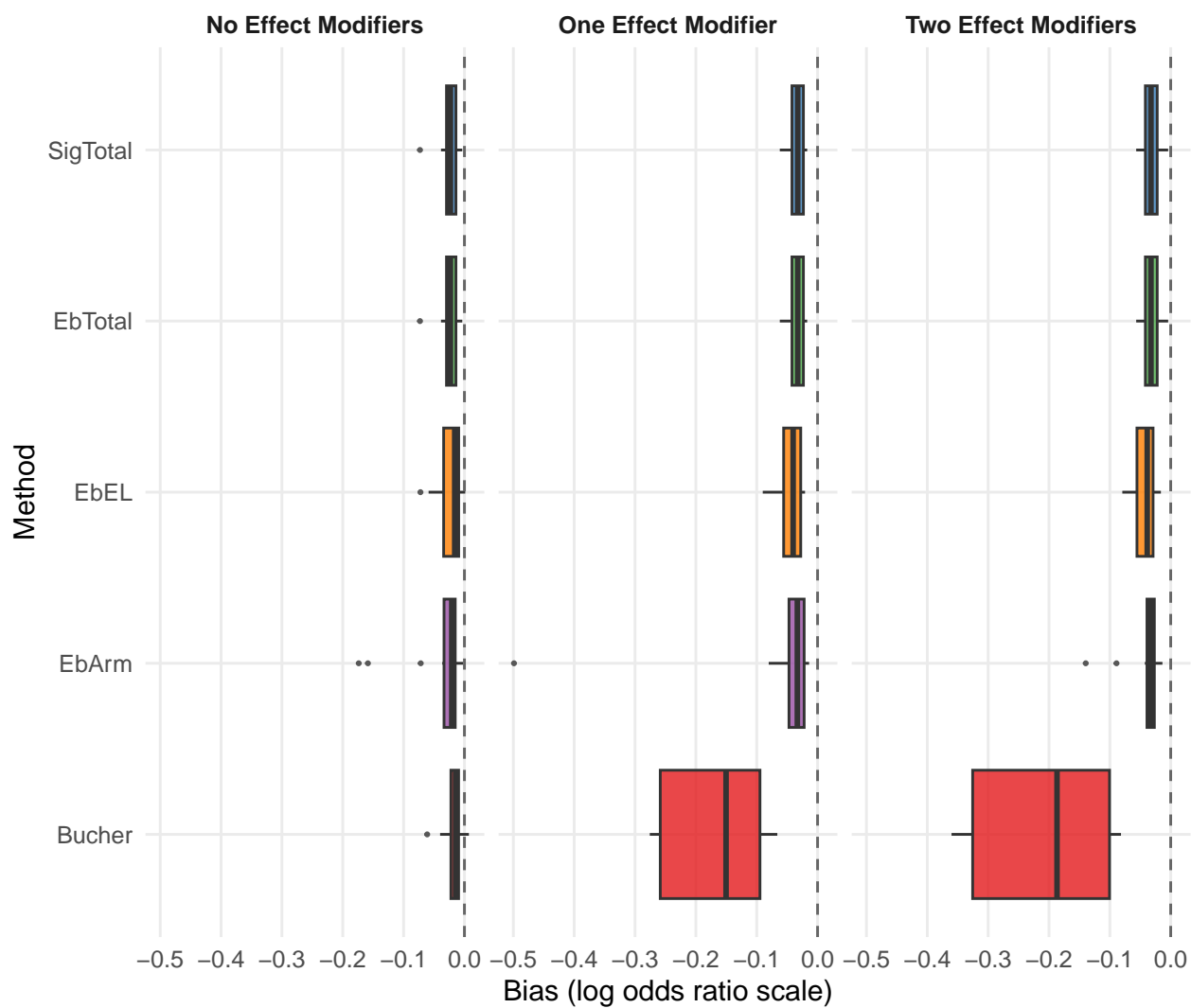


Figure 3: Distribution of bias by number of effect modifiers.

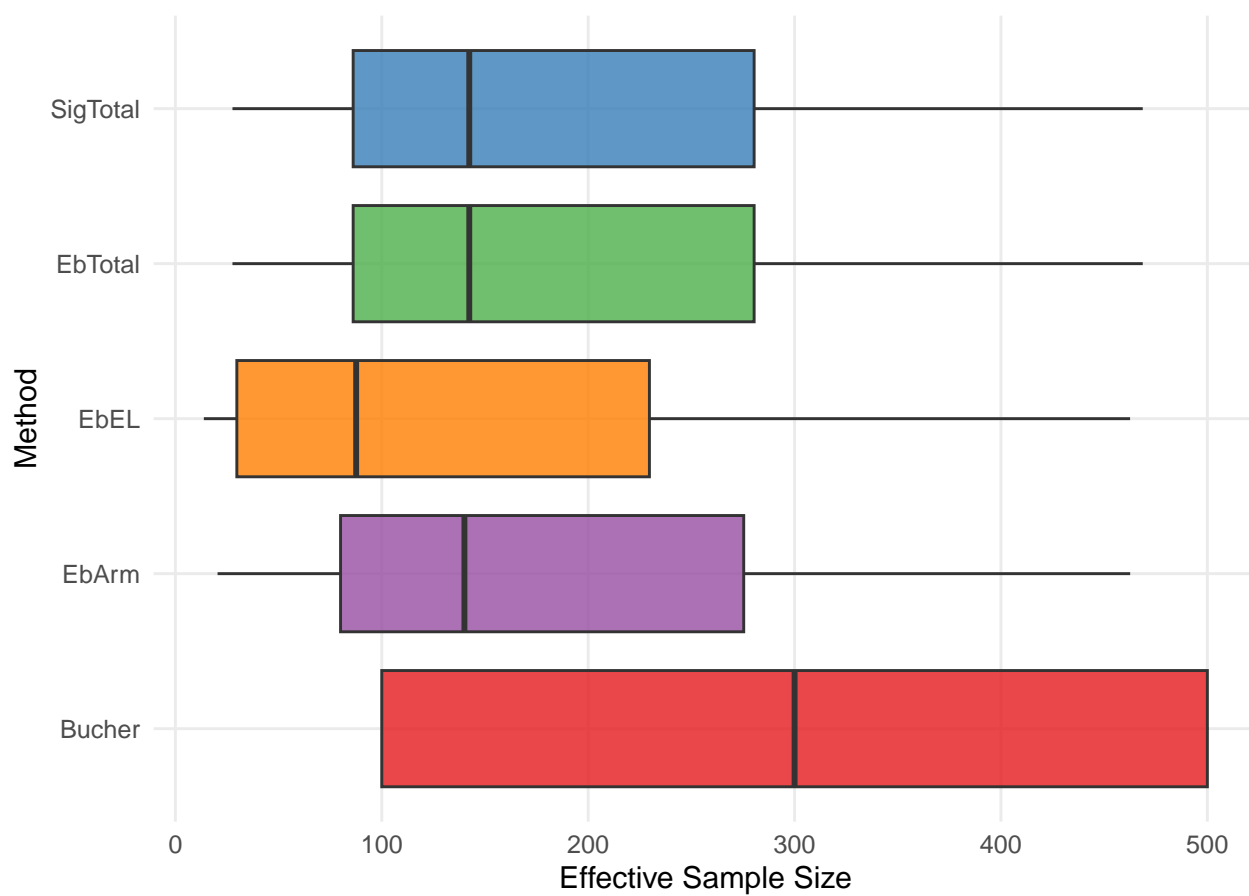


Figure 4: Effective sample size distribution by method. Higher ESS indicates less extreme weighting.

Figure 5 displays the ESS by population overlap, showing the dramatic ESS reduction under low overlap conditions. Under low overlap, EbEL retained only 8% of the original sample size on average (ESS = 24), while EbTotal/SigTotal retained 29% (ESS = 86).

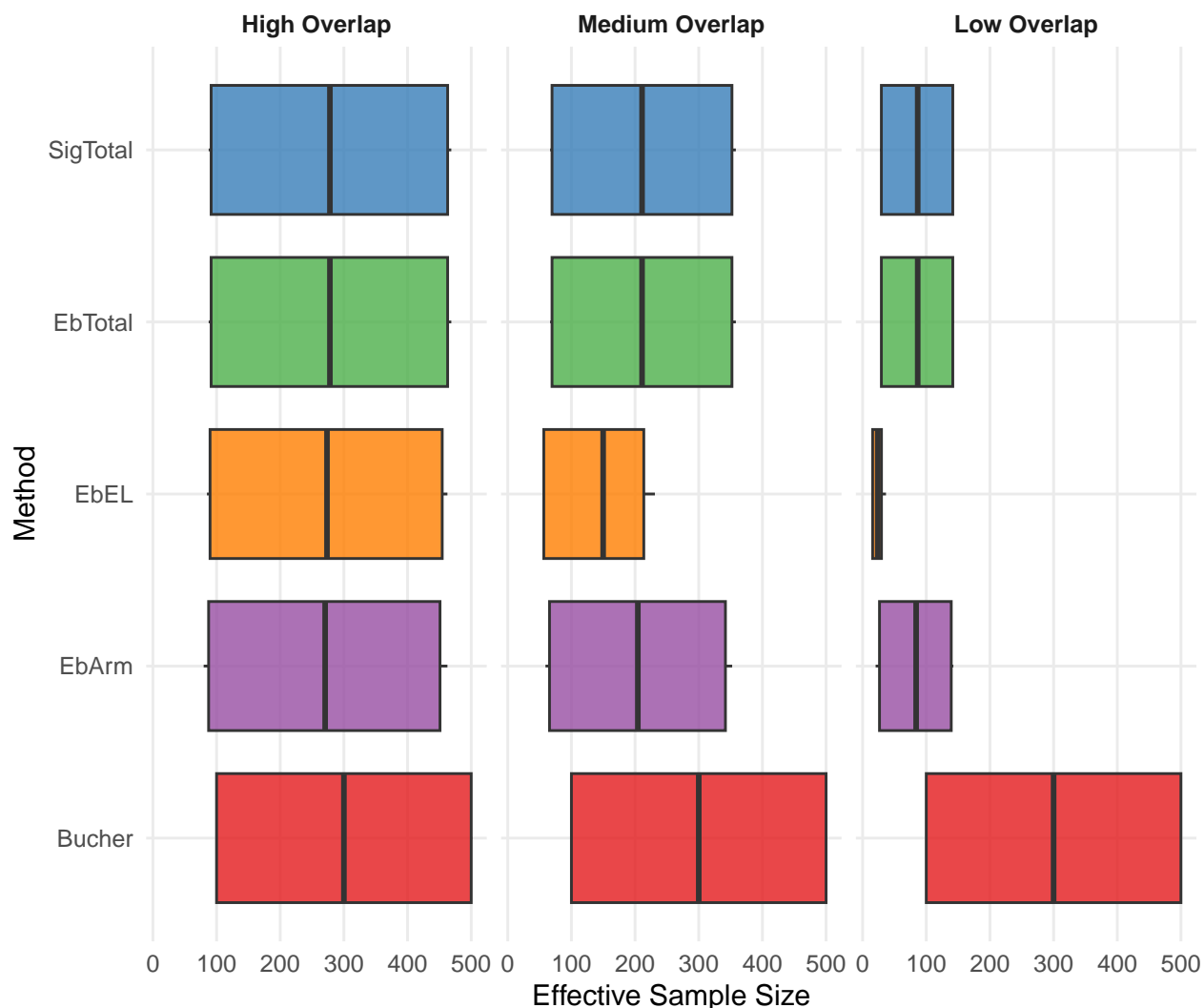


Figure 5: Effective sample size by population overlap level.

### 3.6 Coverage Probability

Figure 6 displays 95% confidence interval coverage by method. EbTotal/SigTotal and EbArm achieved approximately nominal coverage (94.3%). The Bucher method (91.6%) and EbEL (91.8%) showed slight undercoverage, though this may be attributable to bias rather than variance underestimation.

### 3.7 Bias-ESS Tradeoff

Figure 7 illustrates the fundamental tradeoff between bias reduction and effective sample size. The Bucher method achieves the highest ESS (no weighting) but at the cost of substantial bias. MAIC methods reduce bias by weighting, which necessarily reduces ESS. Among the MAIC methods, EbTotal/SigTotal achieved the best balance, with the lowest bias and highest ESS. EbEL showed lower ESS without compensating bias reduction.

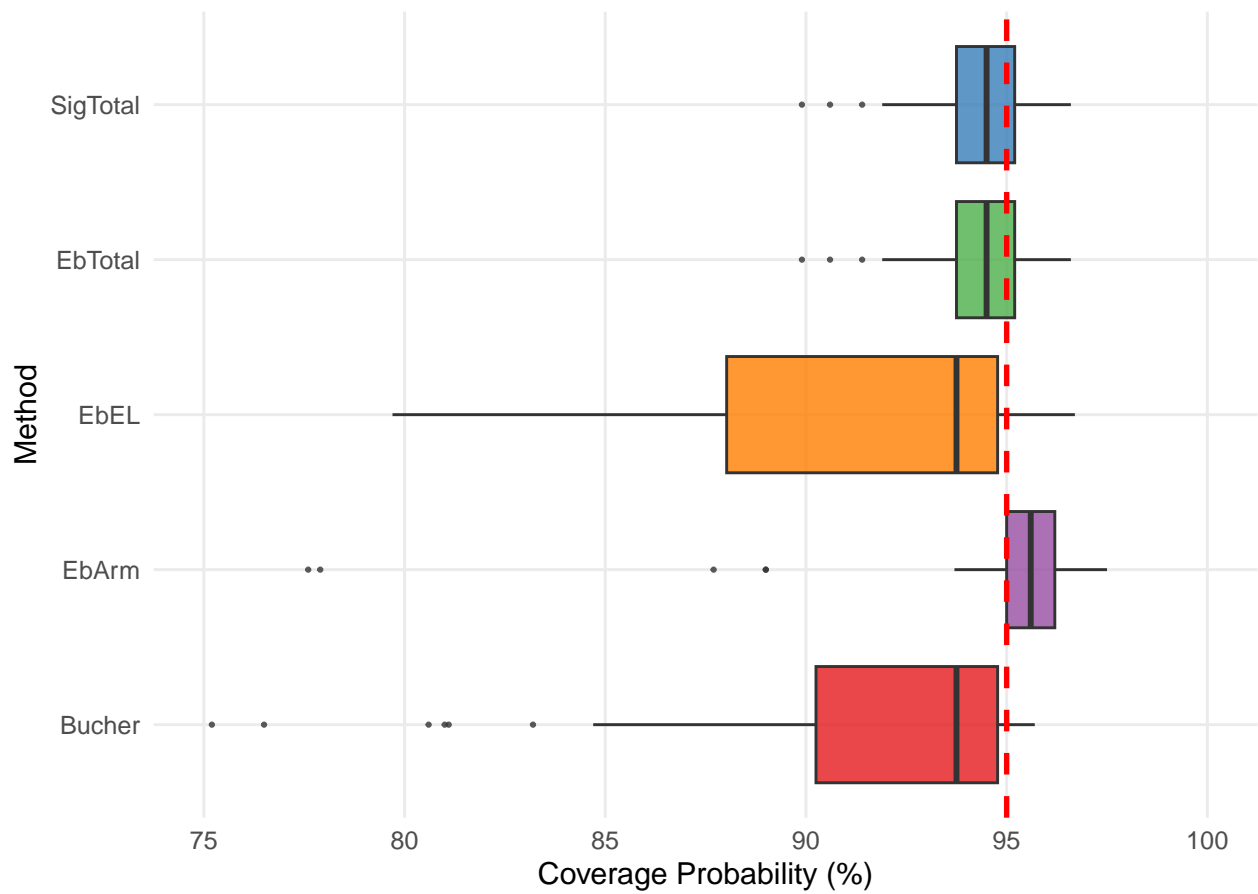


Figure 6: 95% confidence interval coverage by method. The dashed red line indicates nominal 95% coverage.

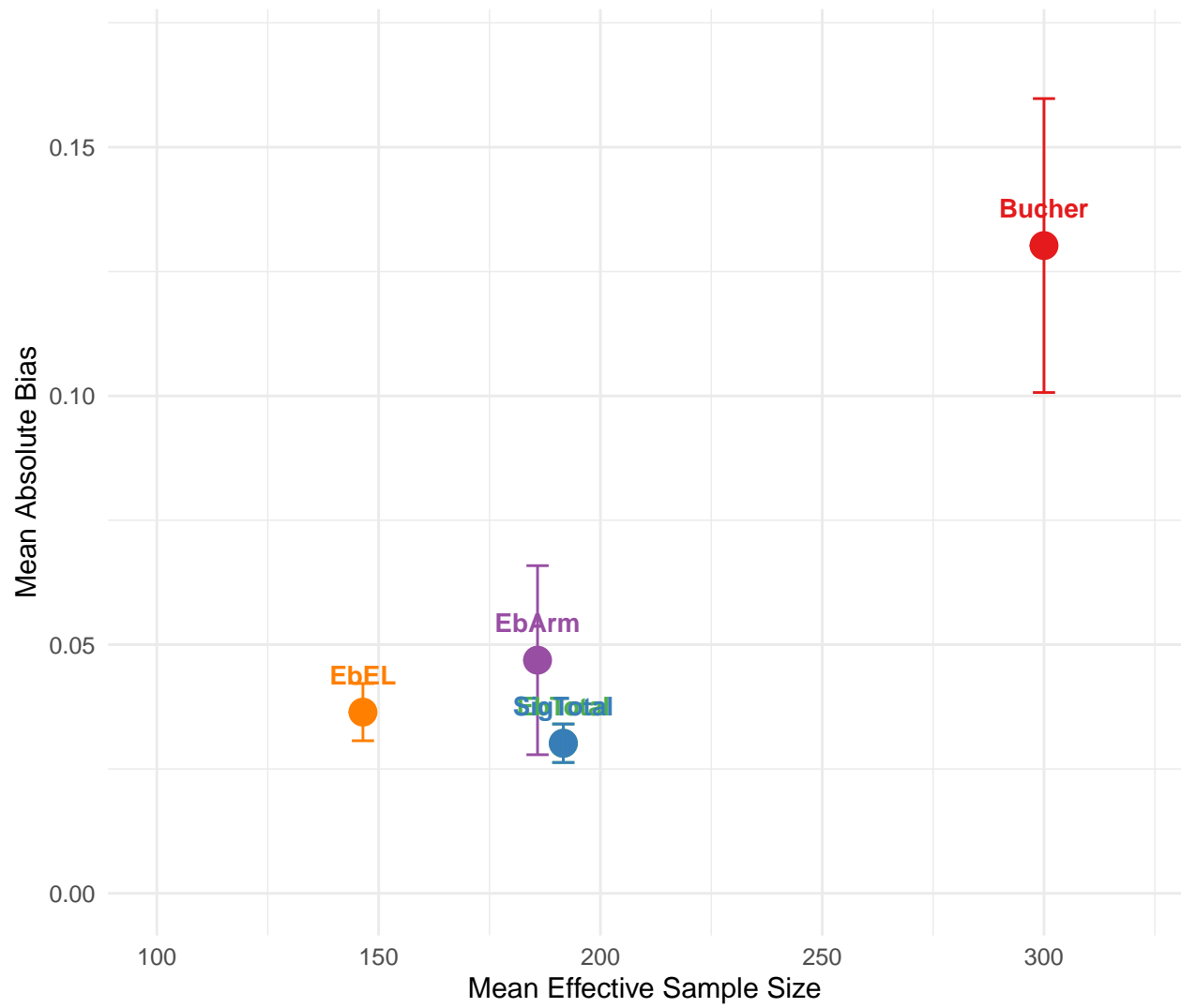


Figure 7: Bias-ESS tradeoff by method. Error bars show 95% confidence intervals.

### 3.8 Scenario-Specific Results

Figure 8 presents a heatmap of mean bias across all scenario configurations. The pattern clearly shows that Bucher method bias increases with decreasing overlap and increasing effect modification, while MAIC methods maintain relatively consistent low bias across conditions. The EbArm method showed elevated bias in some low-overlap scenarios.

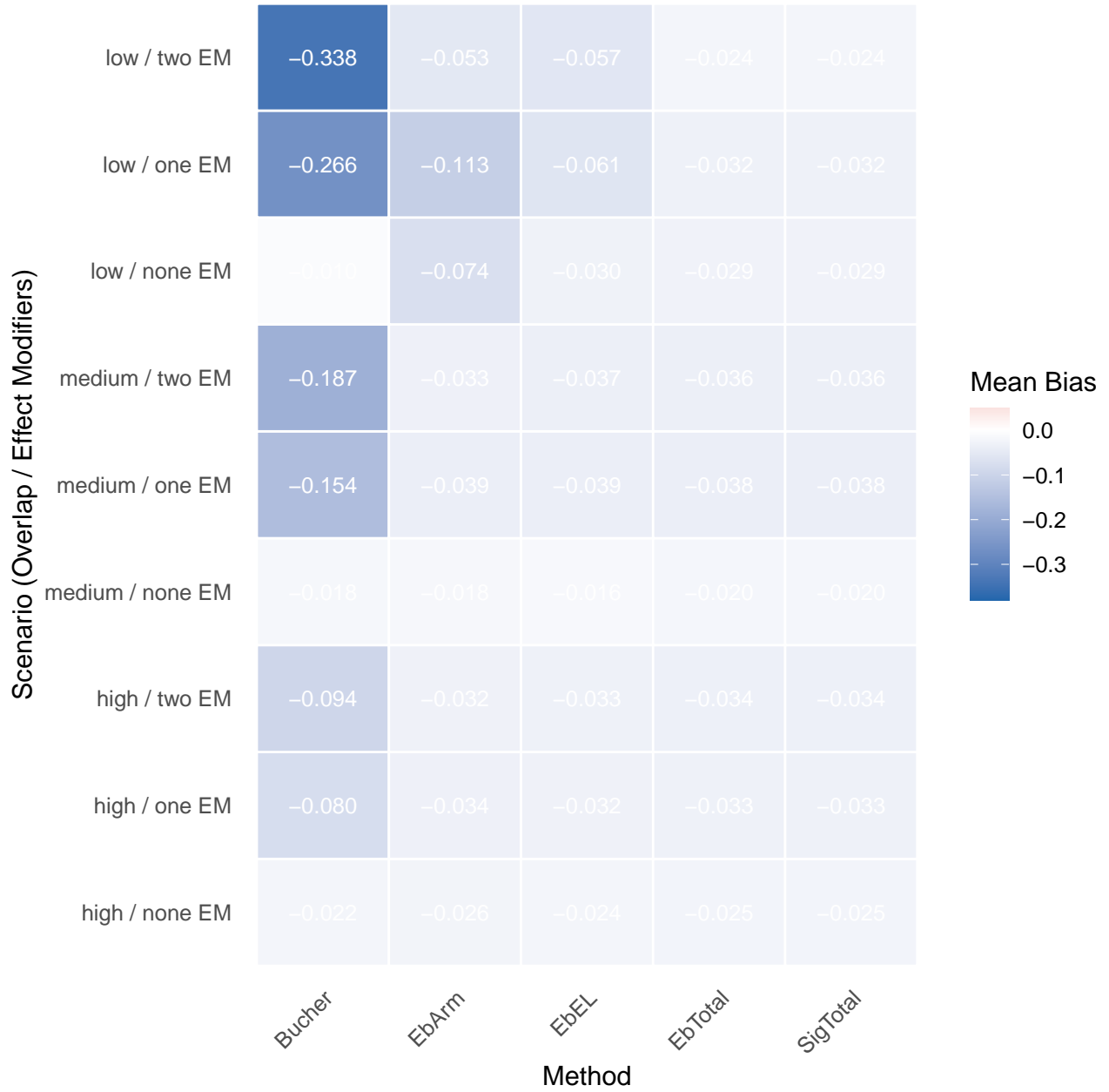


Figure 8: Heatmap of mean bias across scenario configurations. Each cell shows the mean bias for a specific overlap/effect modifier combination.

## 4 Discussion

This comprehensive simulation study provides several key insights for MAIC methodology and practice.

## 4.1 Equivalence of Method of Moments and Entropy Balancing

The empirical confirmation of the theoretical equivalence between method of moments and entropy balancing is a central finding of this study. Across 270,000 simulation iterations, the two approaches produced numerically identical point estimates to within machine precision. This equivalence, first proven theoretically by Phillippo et al. (Phillippo et al. 2020), has important practical implications. Practitioners need not choose between the two approaches on statistical grounds, as they yield identical results. However, the entropy balancing formulation offers conceptual and computational advantages: it naturally extends to non-uniform base weights, clearly specifies the objective being optimized, and connects MAIC to the broader literature on covariate balancing in causal inference.

## 4.2 Value of Population Adjustment When Effect Modification is Present

The simulation results clearly demonstrate that MAIC methods substantially reduce bias when effect modifiers are present and populations differ. The unadjusted Bucher method showed 5 to 10-fold higher bias than MAIC methods when one or two effect modifiers were included in the data-generating model. This finding reinforces the importance of population adjustment in indirect comparisons when cross-trial differences in effect modifiers are suspected.

Conversely, when no effect modifiers were present, MAIC provided minimal bias reduction while increasing variance through weighting. This finding suggests that the decision to apply MAIC should be guided by prior knowledge or evidence regarding effect modification. When effect modification is unlikely, the unadjusted Bucher method may be preferred to avoid unnecessary variance inflation.

## 4.3 Limited Advantage of Alternative Loss Functions

The empirical likelihood method (EbEL), despite its theoretical appeal based on the work of Owen (Owen 2001), did not demonstrate advantages over standard entropy balancing in this simulation. EbEL showed higher RMSE and substantially lower ESS without compensating bias reduction. The lower ESS indicates more extreme weight distributions, which may increase sensitivity to model misspecification and outliers. These results suggest that standard entropy (Kullback-Leibler divergence) remains the preferred loss function for MAIC applications.

## 4.4 Conditional Utility of Arm-Separate Weighting

The arm-separate entropy balancing method (EbArm) showed mixed performance. Under high and medium population overlap, EbArm performed comparably to total-population methods. However, under low overlap, EbArm exhibited substantial instability with RMSE of 2.30, nearly five times higher than EbTotal. This instability likely arises because arm-separate weighting effectively halves the sample size available for each weighting problem, exacerbating the challenges of extreme weighting under low overlap conditions.

These findings suggest that arm-separate weighting should be reserved for scenarios where arm-specific aggregate data are available and population overlap is adequate, ensuring sufficient effective sample size within each arm. A practical guideline might be to require ESS exceeding 50 within each arm before applying arm-separate methods.

## 4.5 Importance of Effective Sample Size Monitoring

Effective sample size emerged as a critical diagnostic for MAIC analyses. Under low overlap conditions, ESS dropped dramatically, with EbEL retaining only 8% of the original sample size on average. Such extreme ESS reduction indicates that only a small fraction of the IPD contributes meaningfully to the weighted analysis, potentially leading to high variance, sensitivity to outliers, and violations of the positivity assumption.

Practitioners should routinely report ESS alongside treatment effect estimates. An ESS below 40% of the original sample size may warrant concern about the reliability of the population adjustment. When ESS is very low, consideration should be given to whether the target population defined by the aggregate data is feasible given the index trial population.



## 4.6 Practical Recommendations

Based on these findings, we offer the following recommendations for MAIC practice. First, entropy balancing (EbTotal) or the equivalent method of moments (SigTotal) should be used as the default approach, as these methods demonstrated the best balance of bias reduction and stability across conditions. Second, the decision to apply MAIC should be informed by prior evidence regarding effect modification; when effect modification is unlikely, unadjusted indirect comparison may be preferred. Third, effective sample size should be monitored as a diagnostic for extreme weighting, with ESS below 40% of the original sample warranting caution. Fourth, arm-separate methods should be reserved for high-overlap scenarios with adequate within-arm ESS. Fifth, alternative loss functions such as empirical likelihood do not appear to offer advantages over standard entropy balancing.

## 4.7 Limitations

Several limitations should be considered when interpreting these results. The simulation used logistic models for binary outcomes; different patterns may emerge with continuous or time-to-event outcomes. The comparator trial sample size was fixed at 500; smaller aggregate data samples may introduce additional variability. The simulation did not evaluate robustness to unmeasured effect modifiers, a fundamental assumption of MAIC that cannot be verified from data. Finally, the simulation used known treatment effects and covariate distributions; real applications face uncertainty about these quantities.

## 5 Conclusions

This simulation study confirms the mathematical equivalence of method of moments and entropy balancing MAIC and provides evidence-based guidance for method selection. Standard entropy balancing is recommended as the first-choice approach for population-adjusted indirect comparisons. Alternative loss functions and arm-separate weighting do not offer clear advantages and may perform worse under challenging conditions. Effective sample size monitoring is essential for identifying potential problems with extreme weighting. Future research should evaluate the robustness of MAIC methods to unmeasured confounding and extend these comparisons to time-to-event outcomes common in oncology applications.

## 6 Data Availability

The advmaic R package implementing all methods, simulation code, and complete results are available at <https://github.com/choxos/advmaic>. Supplementary materials including full scenario-level results tables and additional figures are available in the simulation/results directory of the repository.

## 7 Acknowledgments

The author thanks the developers of the R packages used in this study, particularly dplyr, ggplot2, sandwich, and nloptr.

## References

- Bucher, Heiner C, Gordon H Guyatt, Lauren E Griffith, and Stephen D Walter. 1997. “The Results of Direct and Indirect Treatment Comparisons in Meta-Analysis of Randomized Controlled Trials.” *Journal of Clinical Epidemiology* 50 (6): 683–91.
- Cressie, Noel, and Timothy RC Read. 1984. “Multinomial Goodness-of-Fit Tests.” *Journal of the Royal Statistical Society: Series B (Methodological)* 46 (3): 440–64.
- Owen, Art B. 2001. *Empirical Likelihood*. Chapman; Hall/CRC.

- Petto, Hwee, Zbigniew Kadziola, Alan Brnabic, et al. 2019. “Alternative Weighting Approaches for Anchored Matching-Adjusted Indirect Comparisons via a Common Comparator.” *Value in Health* 22 (1): 85–91.
- Phillippo, David M, Sofia Dias, A E Ades, and Nicky J Welton. 2020. “Equivalence of Entropy Balancing and the Method of Moments for Matching-Adjusted Indirect Comparison.” *Research Synthesis Methods* 11 (4): 568–72.
- Signorovitch, James E, Vanja Sikirica, M Haim Erder, et al. 2012. “Matching-Adjusted Indirect Comparisons: A New Tool for Timely Comparative Effectiveness Research.” *Value in Health* 15 (6): 940–47.
- Signorovitch, James E, Eric Q Wu, Andy P Yu, et al. 2010. “Comparative Effectiveness Without Head-to-Head Trials: A Method for Matching-Adjusted Indirect Comparisons Applied to Psoriasis Treatment with Adalimumab or Etanercept.” *PharmacoEconomics* 28 (10): 935–45.
- Williamson, Elizabeth J, Andrew Forbes, and Ian R White. 2014. “Variance Reduction in Randomised Trials by Inverse Probability Weighting Using the Propensity Score.” *Statistics in Medicine* 33 (5): 721–37.