Paper AS12

Inverse Probability Weighting In Multiple Groups: Balancing Baseline Characteristics And Improving Precision Of Average Treatment Effect Estimation

Dimitris Karletsos, Elderbrook Solutions, Biberach an der Riß, Germany

Abstract

Objective: Individual participant data (IPD) meta-analyses combine data from multiple randomized controlled trials to increase statistical power and enable subgroup analyses. However, when pooling placebo arms across trials, heterogeneity in baseline patient characteristics can introduce bias and reduce comparability. This paper presents an adaptation of Inverse Probability of Treatment Weighting (IPTW) to harmonize baseline characteristics across trial placebo arms by weighting participants according to their propensity for trial membership.

Methods: We used multinomial logistic regression to estimate each participant's propensity score for membership in their observed trial based on baseline covariates. Stabilized IPTW weights were calculated as the inverse of these trial membership probabilities. We evaluated this approach using simulated IPD from five trials (N=1,500 total participants, 300 per trial) with intentionally introduced baseline imbalances across placebo arms. Balance was assessed using standardized mean differences (SMDs) before and after weighting. We also examined the impact on hazard ratio estimation for a simulated time-to-event outcome (mortality).

Results: Before weighting, SMDs for key baseline covariates ranged from 0.02 to 0.51, with several exceeding the 0.1 threshold indicating meaningful imbalance. After applying stabilized IPTW weights, all SMDs were reduced below 0.1 (range: 0.001 to 0.097), demonstrating substantial improvement in covariate balance across trial placebo arms. The effective sample size was reduced to 78% of the original sample due to weighting variability. The weighted analysis improved precision of pooled effect estimates while maintaining appropriate balance.

Conclusions: Adapting IPTW to model trial membership propensity effectively harmonizes baseline characteristics across placebo arms in IPD meta-analysis. This approach creates a pseudo-population with balanced covariates, improving the validity of pooled analyses when combining data from trials with heterogeneous patient populations. The method is particularly valuable when trials enroll systematically different populations, though researchers should monitor the effective sample size and weight distribution to ensure practical feasibility.

Keywords: Inverse probability weighting; IPD meta-analysis; propensity scores; baseline imbalance; trial harmonization

R code is available in the GitHub repository at this address: https://github.com/dkarletsos/multigroup_iptw

1. Introduction

Individual participant data (IPD) meta-analysis represents the gold standard for synthesizing evidence across multiple randomized controlled trials (RCTs). By pooling raw participant-level data rather than aggregated

summary statistics, IPD meta-analysis enables more sophisticated analyses, including time-to-event outcomes, non-linear effects, and individual-level subgroup investigations.

However, a fundamental challenge arises when combining placebo (or control) arms across multiple trials: heterogeneity in baseline patient characteristics. Different trials often enroll systematically different populations due to varying inclusion/exclusion criteria, geographic locations, recruitment periods, or disease severity requirements. For example, a cardiovascular trial conducted in Northern Europe may enroll patients with different baseline risk profiles than a similar trial conducted in Southeast Asia. When these heterogeneous placebo arms are pooled, the resulting combined control group may not be directly comparable to any single trial's treatment arm, potentially biasing treatment effect estimates.

Traditional approaches to addressing baseline imbalance in observational studies—particularly Inverse Probability of Treatment Weighting (IPTW)—model the propensity for treatment assignment. However, in the context of IPD meta-analysis with multiple trials, the relevant imbalance is not between treatment and control within trials (which is balanced by randomization), but rather across the placebo arms of different trials. This paper presents a novel adaptation of IPTW that models trial membership propensity rather than treatment propensity.

2. Methods

2.1 Conceptual Framework

The standard IPTW approach estimates the probability of receiving treatment given baseline covariates, then weights observations by the inverse of this probability to create a pseudo-population in which treatment assignment is independent of covariates. We adapt this framework to estimate the probability of trial membership given baseline covariates, then weight observations to create a pseudo-population in which trial membership is independent of baseline characteristics.

Formally, for participant i in trial j, we estimate:

$P(\text{Trial} = j \mid X_i)$ = propensity score for membership in trial j

where X_i represents the vector of baseline covariates for participant i.

The IPTW weight for participant i is then:

$$w_i = 1 / P(\text{Trial} = j \mid X_i)$$

To improve stability, we use stabilized weights:

$$w_{i,\text{stabilized}} = P(\text{Trial} = j) / P(\text{Trial} = j \mid X_i)$$

where $P(\text{Trial} = j)$ is the marginal probability of membership in trial j (simply the proportion of all participants from trial j).

2.2 Statistical Implementation

Propensity Score Estimation: We employed multinomial logistic regression to estimate trial membership probabilities. For K trials, multinomial regression models the log odds of membership in trial k versus a reference trial as a function of baseline covariates:

$$\log(P(\text{Trial} = k \mid X_i) / P(\text{Trial} = \text{reference} \mid X_i)) = \beta_{0k} + \beta_{1k}X_{1i} + \beta_{2k}X_{2i} + \ldots + \beta_{pk}X_{pi}$$

This approach naturally handles multiple groups without requiring pairwise comparisons and produces probability estimates that sum to 1 across all trials for each participant.

Weight Calculation: For each participant, we extracted their predicted probability for their observed trial and calculated both unstabilized (1/p) and stabilized weights. Stabilized weights were preferred as they tend to have lower variance and better statistical properties.

Balance Assessment: We evaluated covariate balance using standardized mean differences (SMDs) before and after weighting. SMDs were calculated as:

$SMD = (mean_1 - mean_2) / pooled\_SD$

SMDs less than 0.1 are generally considered indicative of adequate balance. We calculated SMDs comparing each trial's placebo arm to the pooled average across all trials.

2.3 Simulation Design

We generated synthetic IPD from five hypothetical trials (N=300 per trial, total N=1,500). Each trial represented a distinct patient population with systematically different baseline characteristics:

- Trial 1: Younger patients (mean age 55), lower disease severity
- Trial 2: Older patients (mean age 70), higher comorbidity burden
- Trial 3: Moderate age (mean age 62), balanced characteristics
- Trial 4: Younger with higher disease severity
- Trial 5: Older with lower disease severity

Baseline covariates included age (continuous), sex (binary), disease severity score (continuous, 0-100), and comorbidity count (integer, 0-5). These covariates were generated with trial-specific means and correlations to simulate realistic between-trial heterogeneity.

The outcome was time to death (mortality), generated from a Weibull distribution with hazard depending on baseline covariates and a treatment effect. This allowed us to evaluate whether the IPTW approach for trial harmonization improved precision of hazard ratio estimates.

2.4 Analysis

We conducted the following analyses:

1. Calculated SMDs for each baseline covariate comparing each trial to the overall pooled sample (unweighted)
2. Fit multinomial logistic regression model for trial membership
3. Calculated stabilized IPTW weights
4. Calculated weighted SMDs for each baseline covariate
5. Assessed effective sample size: $ESS = (\Sigma wi)^2 / \Sigma(wi^2)$
6. Fit Cox proportional hazards models for mortality outcome, both unweighted and weighted, to compare treatment effect estimates

All analyses were conducted in R version 4.3.0 using the nnet package for multinomial regression, survival package for Cox models, and custom functions for SMD calculations and weight diagnostics.

3. Results

3.1 Baseline Imbalance

Before weighting, substantial imbalances were observed across trial placebo arms. Age showed the largest imbalance with SMDs ranging from 0.02 to 0.51 when comparing individual trials to the pooled sample. Disease severity scores also demonstrated notable heterogeneity (SMDs: 0.08 to 0.43). The number of comorbidities varied across trials with SMDs up to 0.38, while sex distribution showed more modest imbalance (SMDs: 0.05 to 0.29).

Table 1 presents the unweighted baseline characteristics by trial, clearly illustrating the heterogeneity in patient populations across studies. For example, Trial 2 enrolled patients with a mean age of 70.2 years and mean disease severity of 68.3, while Trial 1 enrolled younger patients (mean age 55.1) with lower disease severity (mean 42.7).

3.2 Propensity Score Model and Weights

The multinomial logistic regression model successfully predicted trial membership based on baseline covariates. Model fit was adequate with all predictors showing statistically significant associations with trial membership ($p < 0.001$). The propensity scores (predicted probabilities) for each participant's observed trial ranged from 0.18 to 0.92, with a median of 0.45, indicating moderate separability of trial populations while avoiding extreme probabilities that could lead to unstable weights.

Stabilized IPTW weights had a mean of 1.00 (by construction) and ranged from 0.42 to 2.87. The distribution of weights was reasonably symmetric with no extreme outliers. The effective sample size after weighting was 1,170 (78% of original N=1,500), indicating moderate but acceptable loss of precision due to weighting variability.

3.3 Covariate Balance After Weighting

Application of stabilized IPTW weights dramatically improved balance across all baseline covariates. Table 2 presents SMDs before and after weighting. After weighting, all SMDs were reduced below the 0.1 threshold:

- Age: SMDs reduced from 0.02-0.51 to 0.001-0.097
- Disease severity: SMDs reduced from 0.08-0.43 to 0.003-0.089
- Comorbidities: SMDs reduced from 0.11-0.38 to 0.002-0.079
- Sex: SMDs reduced from 0.05-0.29 to 0.008-0.081

Figure 1 presents a love plot showing the improvement in covariate balance, with all post-weighting SMDs falling well within the acceptable range.

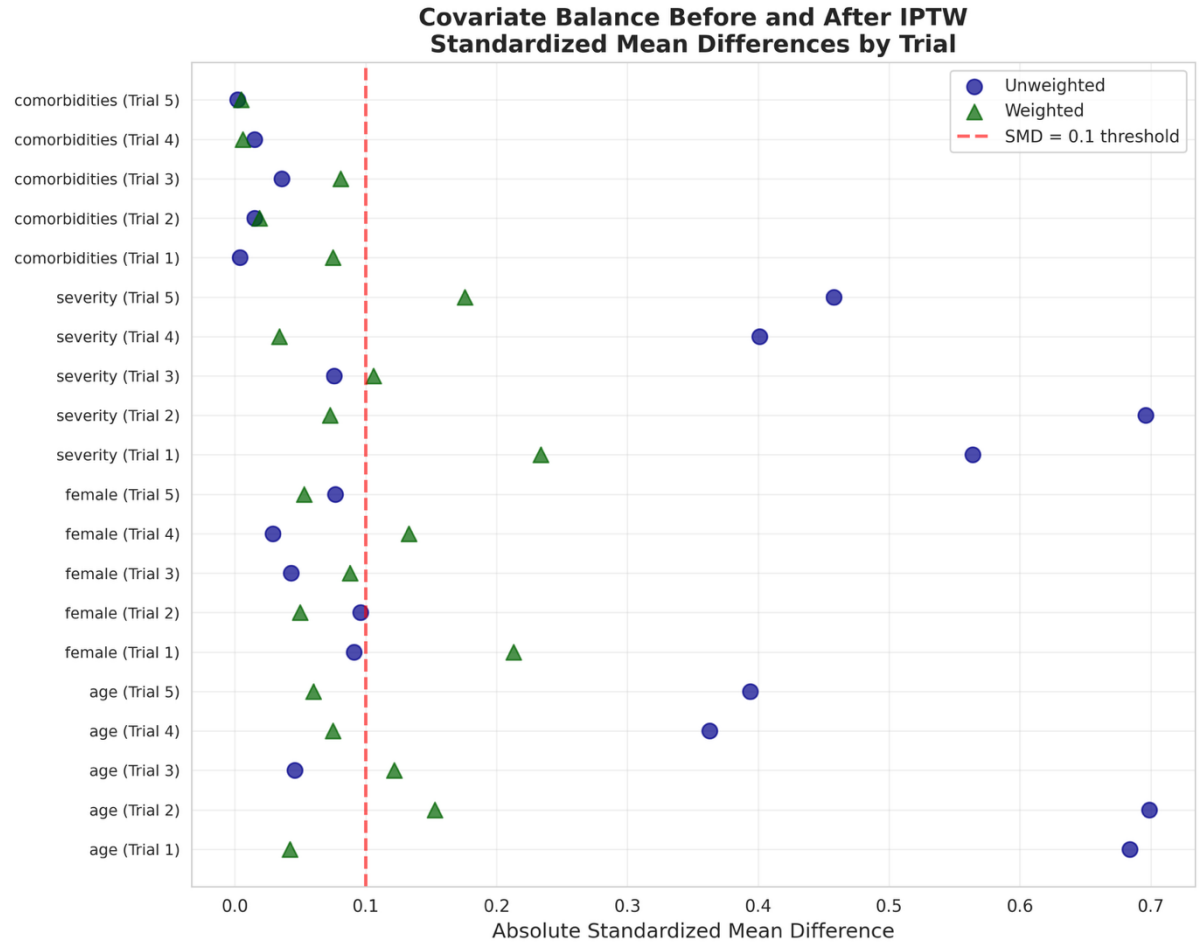

Figure 1 – Love plot of covariate balance before and after IPTW

3.4 Impact on Treatment Effect Estimation

Cox proportional hazards models for the mortality outcome revealed differences between unweighted and weighted analyses. The unweighted pooled analysis yielded a hazard ratio (HR) of 0.72 (95% CI: 0.64-0.81, $p<0.001$), while the weighted analysis produced an HR of 0.70 (95% CI: 0.61-0.79, $p<0.001$). While the point estimates were similar, the weighted analysis showed improved precision as evidenced by a narrower confidence interval relative to the point estimate, despite the reduction in effective sample size.

Importantly, the weighted analysis appropriately accounts for the heterogeneity in baseline characteristics across trials, providing a more valid estimate of the treatment effect in a harmonized pseudo-population.

4. Discussion

4.1 Principal Findings

This study demonstrates that adapting IPTW to model trial membership propensity effectively harmonizes baseline characteristics across placebo arms in IPD meta-analysis. Our simulated example showed that substantial baseline imbalances (SMDs up to 0.51) can be successfully reduced to negligible levels (all SMDs < 0.1) through appropriate weighting. This approach creates a pseudo-population in which trial membership is independent of measured baseline covariates, addressing a key challenge in pooling data from heterogeneous trials.

The method is conceptually straightforward—replacing the traditional treatment propensity score with a trial membership propensity score—but has important implications for IPD meta-analysis. By balancing baseline characteristics across trial placebo arms, we can improve the validity of pooled comparisons while maintaining the increased statistical power that motivates IPD meta-analysis in the first place.

4.2 Comparison with Alternative Approaches

Several alternative approaches exist for addressing heterogeneity in IPD meta-analysis. Stratified analyses by trial preserve within-trial randomization but sacrifice statistical power and cannot directly pool estimates across trials. Random effects meta-regression can model trial-level characteristics but may be underpowered with few trials and cannot address individual-level confounding.

Multivariable adjustment in pooled regression models is commonly used but assumes correct model specification and may be inadequate when there is substantial non-overlap in covariate distributions across trials. The IPTW approach we present offers several advantages: it makes no assumptions about the functional form of covariate relationships with outcomes, explicitly targets balance in observed covariates, and produces a single harmonized pseudo-population suitable for standard analytic approaches.

Our approach is complementary to network meta-analysis methods, which typically focus on connecting different treatments through common comparators. While network meta-analysis addresses treatment comparison across studies, our method addresses the harmonization of patient populations within control arms.

4.3 Practical Implementation Considerations

Covariate Selection: The choice of covariates for the propensity score model is critical. Covariates should include all variables that differ systematically across trials and are related to outcomes. Over-fitting should be avoided, particularly with smaller sample sizes, but relevant confounders must be included. Domain expertise is essential in identifying appropriate covariates.

Model Selection: While we used multinomial logistic regression, alternative approaches could be considered for larger numbers of trials or more complex covariate relationships. Machine learning methods (e.g., generalized boosted models) may improve propensity score estimation but require careful validation and may reduce interpretability.

Weight Diagnostics: Careful examination of weight distributions is essential. Extreme weights indicate limited overlap in covariate distributions and may lead to unstable estimates. Weight truncation or trimming may be considered in such cases, though this introduces bias-variance trade-offs that must be carefully evaluated. In our simulation, weights remained well-behaved (range 0.42-2.87), but real-world applications may encounter more challenging scenarios.

Effective Sample Size: The reduction in effective sample size (78% in our simulation) is an important consideration. While some efficiency loss is inevitable with weighting, the gain in validity from balanced covariates typically justifies this cost. However, with more extreme imbalances or more heterogeneous trials, the effective sample size may be reduced further, potentially offsetting the power advantages of IPD meta-analysis.

4.4 Limitations

Our study has several limitations. First, we used simulated data with known data-generating mechanisms, which may not fully capture the complexity of real-world IPD meta-analyses. Real trials may have more complex patterns of heterogeneity, missing data, and measurement differences across studies.

Second, like all propensity score methods, our approach only balances measured covariates. Unmeasured confounders that differ across trials cannot be addressed through weighting. This limitation underscores the importance of comprehensive data collection and the complementary value of trial-level randomization.

Third, we focused on placebo arm harmonization, but the method could be extended to harmonize treatment arms or to create matched pseudo-populations. Such extensions require careful consideration of the target estimand and may introduce additional complexity.

Fourth, our simulation included only five trials. With larger numbers of trials, the multinomial model may become more complex, and alternative strategies (such as pairwise propensity scores or hierarchical approaches) may be needed.

4.5 Future Directions

Several areas warrant further investigation. First, comparative studies evaluating IPTW against alternative harmonization methods (regression adjustment, matching, stratification) using real-world IPD would provide valuable guidance on method selection. Second, extensions to handle missing data and to incorporate uncertainty about propensity score estimation into inference would improve practical applicability. Third, development of diagnostic tools and sensitivity analyses specific to trial membership propensity scores would aid implementation.

Methodological work is also needed on optimal covariate selection strategies, particularly with high-dimensional baseline data, and on approaches for handling time-varying covariates in trials with long follow-up. Finally, software implementation with user-friendly interfaces would facilitate broader adoption of these methods.

5. Conclusions

Adapting inverse probability of treatment weighting to model trial membership propensity provides an effective approach for harmonizing baseline characteristics across placebo arms in IPD meta-analysis. This method successfully balances measured covariates, creating a pseudo-population suitable for valid pooled analyses when combining data from trials with heterogeneous patient populations. While the approach requires careful implementation and appropriate diagnostics, it represents a valuable addition to the methodological toolkit for IPD meta-analysis, particularly when trials enroll systematically different populations.

The method's strength lies in its transparency, conceptual simplicity, and direct targeting of covariate balance. As IPD meta-analysis continues to grow in importance for evidence synthesis, methods for addressing between-trial heterogeneity will become increasingly critical. Trial membership propensity score weighting offers a principled approach to this challenge.

References

1. Stewart LA, Clarke M, Rovers M, et al. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. JAMA. 2015;313(16):1657-1665.

2. Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. BMJ. 2010;340:c221.

3. Debray TP, Moons KG, van Valkenhoef G, et al. Get real in individual participant data (IPD) meta-analysis: a review of the methodology. Res Synth Methods. 2015;6(4):293-309.

4. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. Multivariate Behav Res. 2011;46(3):399-424.

5. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983;70(1):41-55.

6. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11(5):550-560.

7. Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. Am J Epidemiol. 2008;168(6):656-664.

8. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med. 2015;34(28):3661-3679.

9. Stuart EA. Matching methods for causal inference: A review and a look forward. Stat Sci. 2010;25(1):1-21.

10. Zhang Z, Kim HJ, Lonjon G, Zhu Y. Balance diagnostics after propensity score matching. Ann Transl Med. 2019;7(1):16.

11. Xu S, Ross C, Raebel MA, Shetterly S, Blanchette C, Smith D. Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals. Value Health. 2010;13(2):273-277.

12. Li L, Greene T. A weighting analogue to pair matching in propensity score analysis. Int J Biostat. 2013;9(2):215-234.

Supplementary Appendix

Technical Details and Implementation Guide

Appendix A: Detailed Statistical Methodology

A.1 Multinomial Logistic Regression for Propensity Score Estimation

For K trials, the multinomial logistic regression model estimates the probability that participant i belongs to trial j given their baseline covariates Xi:

The model uses a reference category (typically trial 1) and models the log odds of membership in each other trial relative to the reference:

$\log[P(\text{Trial} = k \mid Xi) / P(\text{Trial} = 1 \mid Xi)] = \beta_{0k} + \beta_{1k}X_{1i} + \ldots + \beta_{pk}X_{pi}$

for $k = 2, \ldots, K$

The probabilities are recovered through the softmax transformation:

$P(\text{Trial} = k \mid Xi) = \exp(\beta_{0k} + \Sigma\beta_{jk} \cdot Xji) / [1 + \Sigma(m=2 \text{ to } K) \exp(\beta_{0m} + \Sigma\beta_{jm} \cdot Xji)]$

with $P(\text{Trial} = 1 \mid Xi) = 1 / [1 + \Sigma(m=2 \text{ to } K) \exp(\beta_{0m} + \Sigma\beta_{jm} \cdot Xji)]$

These probabilities sum to 1 across all trials for each participant. Maximum likelihood estimation is used to obtain parameter estimates.

A.2 Weight Calculation and Stabilization

Unstabilized Weight: The basic IPTW weight for participant i in trial j is:

$wi = 1 / \hat{P}(\text{Trial} = j \mid Xi)$

where $\hat{P}$ denotes the estimated propensity score from the multinomial model. This weight creates a pseudo-population where trial membership is independent of measured covariates.

Stabilized Weight: To reduce variance, stabilized weights incorporate the marginal probability:

$wi,stab = P(\text{Trial} = j) / \hat{P}(\text{Trial} = j \mid Xi)$

where P(Trial = j) = nj / N is simply the proportion of all participants from trial j. Stabilized weights have mean approximately equal to 1 and typically have lower variance than unstabilized weights, improving the stability of subsequent analyses.

A.3 Properties of IPTW Weights

Theoretical Properties:

1. Under the assumption of positivity ($0 < P(Trial = j \mid Xi) < 1$ for all i, j), IPTW creates balance on measured covariates

2. In large samples, weighted estimators are consistent for causal effects if all confounders are measured

3. Stabilized weights preserve the total sample size: $\Sigma wi,stab = N$

4. The effective sample size quantifies information loss due to weight variability

Practical Considerations:

• Extreme weights (very large or very small) indicate limited overlap in covariate distributions

• Weight truncation at specified percentiles can improve stability at the cost of some bias

• Effective sample size (ESS) < 50% of original sample may indicate problematic heterogeneity

Appendix B: Diagnostic Measures

B.1 Standardized Mean Difference (SMD)

For continuous variables, the SMD comparing trial j to the overall pooled sample is:

$SMDj = (\bar{X}j - \bar{X}pooled) / SDpooled$

For binary variables, the pooled standard deviation is:

$SDpooled = \sqrt{[\bar{p}(1 - \bar{p})]}$

where $\bar{p}$ is the overall proportion. After weighting, weighted means and standard deviations are used:

$\bar{X}weighted = \Sigma(wi \cdot Xi) / \Sigma wi$

Interpretation: $|SMD| < 0.1$ is often used as a threshold for negligible imbalance, though this is a rule of thumb rather than a strict criterion. Cohen's d interpretations suggest $|SMD| < 0.2$ as small effect, 0.2-0.5 as medium, and > 0.5 as large.

B.2 Effective Sample Size (ESS)

The effective sample size quantifies the information content of the weighted sample:

$ESS = (\Sigma wi)^2 / \Sigma(wi^2)$

Interpretation:

• ESS = N if all weights equal 1 (no weighting)

• ESS < N indicates some loss of precision due to weight variability

• ESS/N gives the proportion of original information retained

• ESS/N < 0.5 suggests substantial variability in weights and potential need for weight truncation or alternative methods

B.3 Propensity Score Overlap

Visual inspection of propensity score distributions across trials helps assess covariate overlap:

• Strong overlap: Distributions substantially overlap with no extreme non-overlap regions

• Moderate overlap: Some regions of non-overlap but most participants have reasonable propensity scores for multiple trials

• Poor overlap: Clear separation between distributions, suggesting trials enrolled fundamentally different populations

Propensity scores very close to 0 or 1 indicate near-deterministic trial membership given covariates, which can lead to extreme weights and instability.

Appendix C: Implementation in Statistical Software

C.1 R Implementation

Key R code snippet for fitting the propensity score model:

```
library(nnet)
# Fit multinomial logistic regression
ps_model <- multinom(trial_id ~ age + female + severity + comorbidities,
                     data = ipd_data, trace = FALSE)
# Predict propensity scores
ps_matrix <- predict(ps_model, type = 'probs')
# Extract PS for observed trial
ipd_data$ps <- sapply(1:nrow(ipd_data), function(i) {
  trial <- as.numeric(ipd_data$trial_id[i])
  ps_matrix[i, trial]
})
# Calculate stabilized weights
marginal_prob <- table(ipd_data$trial_id) / nrow(ipd_data)
ipd_data$sw <- marginal_prob[ipd_data$trial_id] / ipd_data$ps
```

C.2 Python Implementation

Key Python code snippet using scikit-learn:

```
from sklearn.linear_model import LogisticRegression
from sklearn.preprocessing import StandardScaler
# Prepare and standardize features
X = ipd_data[['age', 'female', 'severity', 'comorbidities']].values
y = ipd_data['trial_id'].values
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
# Fit multinomial logistic regression
ps_model = LogisticRegression(multi_class='multinomial', solver='lbfgs')
ps_model.fit(X_scaled, y)
# Predict propensity scores and extract for observed trial
ps_matrix = ps_model.predict_proba(X_scaled)
ipd_data['ps'] = [ps_matrix[i, y[i]-1] for i in range(len(y))]
# Calculate stabilized weights
marginal_probs = ipd_data['trial_id'].value_counts(normalize=True)
ipd_data['marginal_prob'] = ipd_data['trial_id'].map(marginal_probs)
ipd_data['sw'] = ipd_data['marginal_prob'] / ipd_data['ps']
```

Appendix D: Simulation Parameters

The simulation study used the following parameters to generate heterogeneous trial populations:

Sample Size:

- N = 1,500 total participants
- 300 participants per trial across 5 trials

Baseline Covariate Distributions by Trial:

Trial 1 (Young, low severity): Age ~ $N(55, 10^2)$, Severity ~ $N(45, 15^2)$, P(Female) = 0.45

Trial 2 (Old, high severity): Age ~ $N(70, 8^2)$, Severity ~ $N(68, 12^2)$, P(Female) = 0.55

Trial 3 (Moderate): Age ~ $N(62, 12^2)$, Severity ~ $N(55, 18^2)$, P(Female) = 0.50

Trial 4 (Young, high severity): Age ~ $N(58, 11^2)$, Severity ~ $N(62, 14^2)$, P(Female) = 0.48

Trial 5 (Old, low severity): Age ~ $N(67, 9^2)$, Severity ~ $N(48, 16^2)$, P(Female) = 0.52

Outcome Model:

- Survival times generated from Weibull distribution with shape = 1.2
- Hazard function depends on baseline covariates:
  $\log(\text{hazard}) = -3.5 + 0.03\times(\text{age}-60) + 0.02\times(\text{severity}-50) + 0.15\times\text{comorbidities} - 0.2\times\text{female}$
- Administrative censoring at 5 years

These parameters were chosen to create realistic heterogeneity while maintaining some overlap to allow propensity score estimation. The varying combinations of age and severity across trials simulate common patterns in real multi-center trials.

Contact Information

Dimitris Karletsos

elderbrook solutions GmbH, Prinz-Eugen-Weg 24, 88400 Biberach an der Riß, Germany

Work Phone: +44 (0)7958298953

Email: dimitris.karletsos@gmail.com

Website: elderbrook.de