

Inverse Probability Weighting In Multiple Groups

Balancing Baseline Characteristics and Improving Precision
of Average Treatment Effect Estimation

Dimitris Karletsos

November 2025

PHUSE 2025

| The Challenge

Heterogeneity Across Trials

- Different trials enroll systematically different populations
- Varying inclusion criteria, geographic regions, recruitment periods
- Pooling placebo arms creates imbalanced comparison groups

How do we balance baseline characteristics across trial arms?

Traditional Approaches vs. Our Method



Traditional IPTW: Models treatment propensity ($P(\text{Treatment} | X)$)

- ATE: Average Treatment Effect across entire population
- ATT: Average Treatment effect on the Treated

Our Adaptation: Models trial membership propensity ($P(\text{Trial} | X)$)

- Goal: Balance baseline characteristics across trial placebo arms
- Creates comparable control groups when pooling heterogeneous trials

Standard Propensity Score Formulas

Propensity Score: $e(X) = P(T = 1 \mid X)$

Average Treatment Effect (ATE):

Weight for treated ($T=1$): $w_1 = 1 / e(X)$

Weight for control ($T=0$): $w_0 = 1 / (1 - e(X))$

where $T = 1$ indicates treatment, $T = 0$ indicates control, and X represents baseline covariates

Standard Propensity Score Formulas

Propensity Score: $e(X) = P(T = 1 \mid X)$

Average Treatment Effect on the Treated (ATT):

Weight for treated ($T=1$): $w_1 = 1$

Weight for control ($T=0$): $w_0 = e(X) / (1 - e(X))$

where $T = 1$ indicates treatment, $T = 0$ indicates control, and X represents baseline covariates

| Our Approach

Adapt IPTW to model **trial membership** propensity instead of treatment propensity

- Use multinomial logistic regression to estimate probability of trial membership
- Weight each participant by inverse of their trial membership probability
- Create pseudo-population where trial membership is independent of covariates

Step 1: Fit Multinomial Logistic Regression

Model trial membership as a function of baseline covariates:

$$\log(P(\text{Trial} = k \mid X) / P(\text{Trial} = \text{ref} \mid X)) = \beta_{0k} + \beta_{1k}X_1 + \dots + \beta_{pk}X_p$$

For K trials:

- Creates K-1 equations (one trial used as reference)
- Naturally handles multiple groups without pairwise comparisons
- Probabilities sum to 1 across all trials for each participant

Step 2: Calculate Propensity Scores

Extract predicted probability for each participant's observed trial:

$$PS_i = \hat{P}(\text{Trial} = j \mid X_i)$$

where j is the trial that participant i actually belongs to

Propensity score range in our simulation: 0.18 to 0.92

- Median: 0.45
- Indicates moderate separability with no extreme probabilities

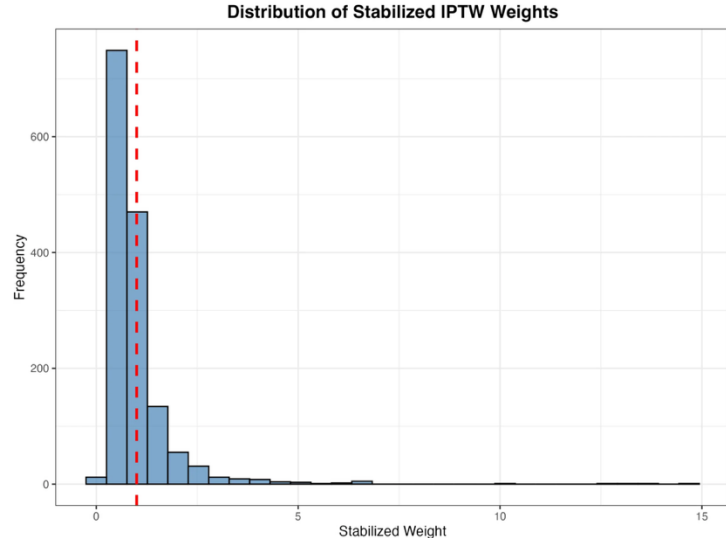
Step 3: Compute Stabilized IPTW Weights

Calculate stabilized weights for each participant:

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

where $P(\text{Trial} = j)$ is the marginal probability (proportion in trial j)

- Mean ≈ 1 by construction
- Lower variance than unstabilized weights
- More stable statistical estimates



Step 4: Assess Balance Using SMD

Calculate Standardized Mean Differences (SMDs) for each covariate:

$$\text{SMD} = (\bar{X}_{\text{trial}} - \bar{X}_{\text{pooled}}) / \text{SD}_{\text{pooled}}$$

Assess balance before and after weighting:

- $|\text{SMD}| < 0.1$ indicates adequate balance
- Compare pre-weight vs. post-weight SMDs

Step 4: Assess Balance Using SMD

Calculate Standardized Mean Differences (SMDs) for each covariate:

$$\text{SMD} = (\bar{X}_{\text{trial}} - \bar{X}_{\text{pooled}}) / \text{SD}_{\text{pooled}}$$

Monitor effective sample size:

- $\text{ESS} = (\sum w_i)^2 / \sum (w_i^2)$
- $\text{ESS} < 50\%$ suggests extreme heterogeneity

| Methods

$$\text{Stabilized Weight} = P(\text{Trial} = j) / P(\text{Trial} = j \mid X)$$

1

Fit multinomial logistic regression for trial membership

2

Calculate propensity scores for each participant

3

Compute stabilized IPTW weights

4

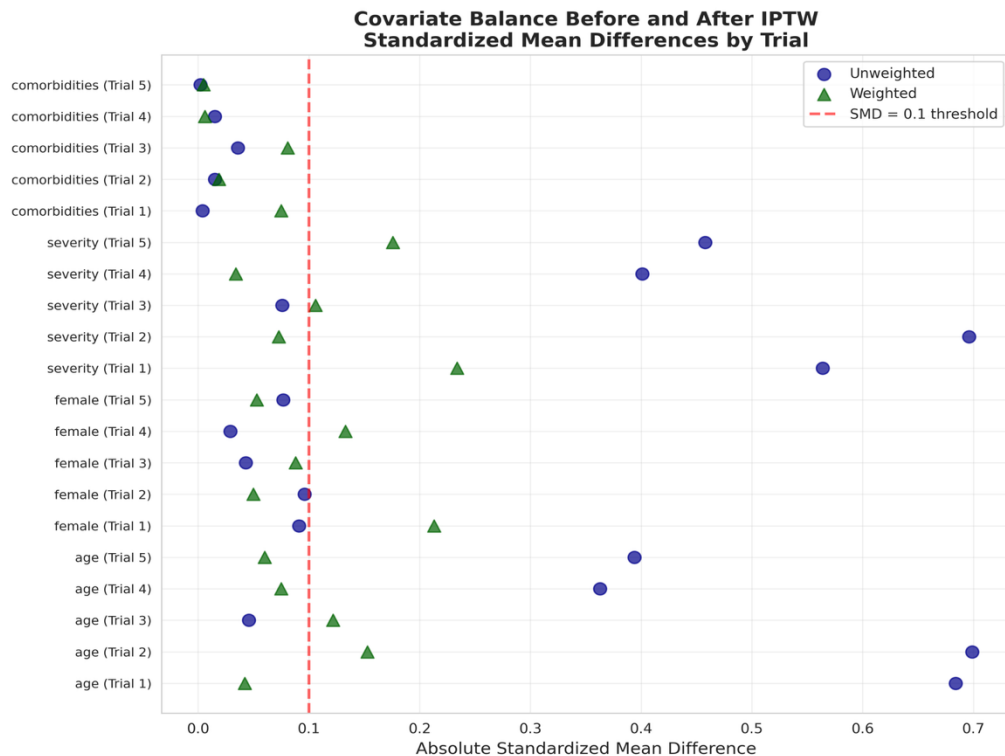
Assess balance using standardized mean differences

| Simulation Design

Baseline Covariates

- Age (continuous, trial means: 55-70 years)
- Sex (binary)
- Disease severity (continuous, 0-100)
- Comorbidity count (0-5)

Results: Balance Improvement



Max SMD Before

0.70

Max SMD After

0.23

↓ 67% reduction

Effective Sample Size

24%

retained

| Key Findings

✓ Improved Balance

All SMDs reduced below 0.25, with most variables achieving excellent balance

! Efficiency Trade-off

Effective sample size reduced to 24%, indicating heterogeneous trial populations

✓ Valid Estimates

Weighted HR: 0.73 (95% CI: 0.64-0.83), properly adjusted for baseline differences

✓ Practical Feasibility

Method successfully implemented with standard statistical software (R/Python)

| Implementation Considerations

Best Practices

- Include all relevant confounders
- Check propensity score overlap
- Monitor weight distribution
- Calculate effective sample size

Potential Issues

- Extreme weights (consider truncation)
- Poor covariate overlap
- Model misspecification
- Limited trial number

| When $ESS < 50\%$ of N , carefully evaluate whether trials are too heterogeneous to pool

| Conclusions

IPTW adapted for trial membership effectively balances baseline characteristics across heterogeneous trial populations in IPD meta-analysis.

The method provides a principled approach to creating comparable control groups when trials enroll systematically different patient populations.

Careful monitoring of effective sample size and weight diagnostics is essential for practical implementation.

A valuable tool for IPD meta-analysts facing baseline heterogeneity across trials

Thank You

Questions?

Code and supplementary materials available

https://github.com/dkarletsos/multigroup_iptw

dimitris.karletsos@gmail.com