COMPARISON OF ESTIMATION METHODS FOR SINGLE-ARM TRIALS IN RARE DISEASES WITH HISTORICAL CONTROL GROUPS

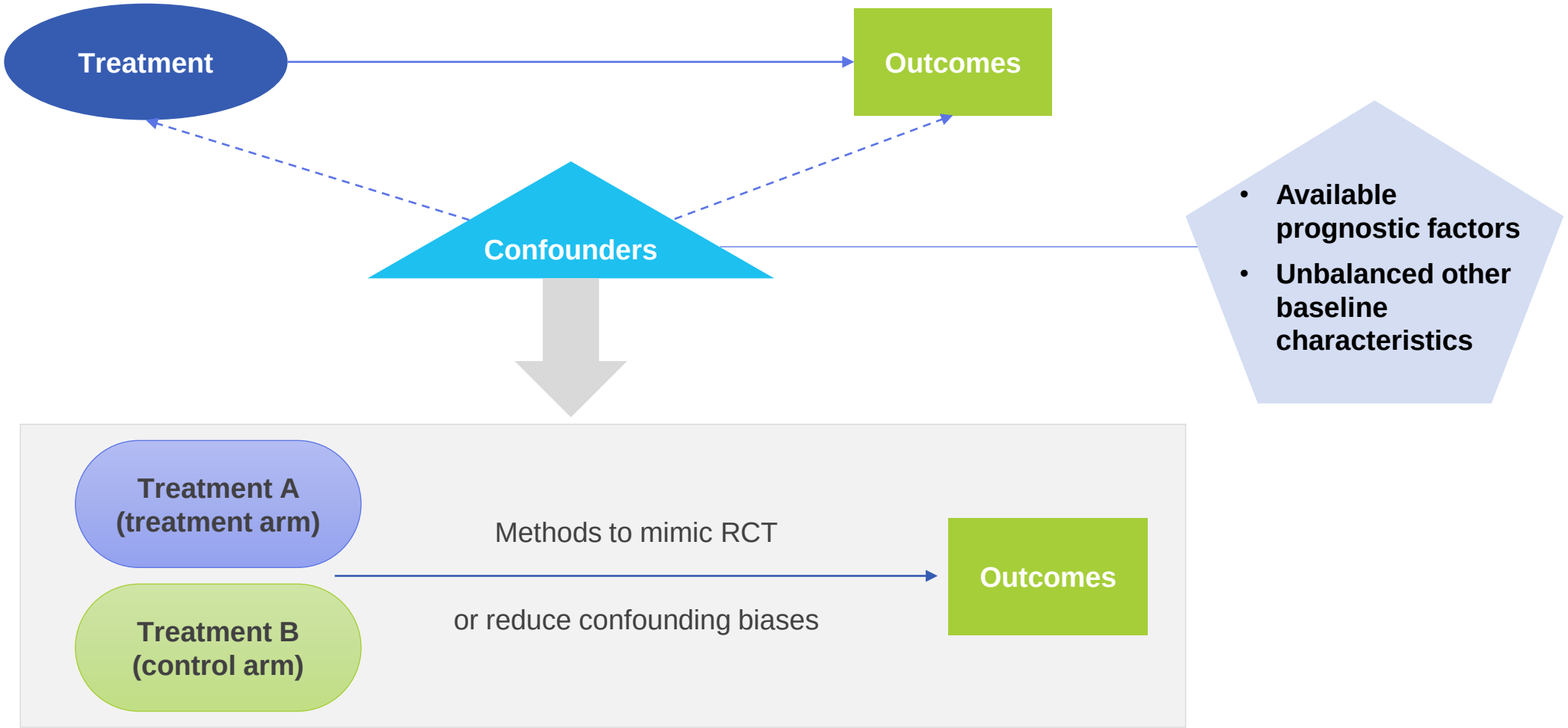
Barlev A[1], Brookhart MA[2], Xun P[3], Thirumalai D[3], Sadetsky N[1], Suissa S[4]

[1]Atara Biotherapeutics, Inc., South San Francisco, CA, USA, [2]Duke University, Durham, NC, USA, [3]Atara Biotherapeutics, Inc., Thousand Oaks, CA, USA, [4]Mc Gill University, Montreal, Quebec, Canada

Conflict of Interest

- A Barlev, P Xun, D Thirumalai and N Sadetsky are employees and shareholders of Atara Biotherapeutics.
- MA Brookhart and S Suissa participated in scientific advisory committee for Atara
- The studies were funded by Atara Biotherapeutics.

There Are Challenges Associated With Deriving Evidence From Real World Data

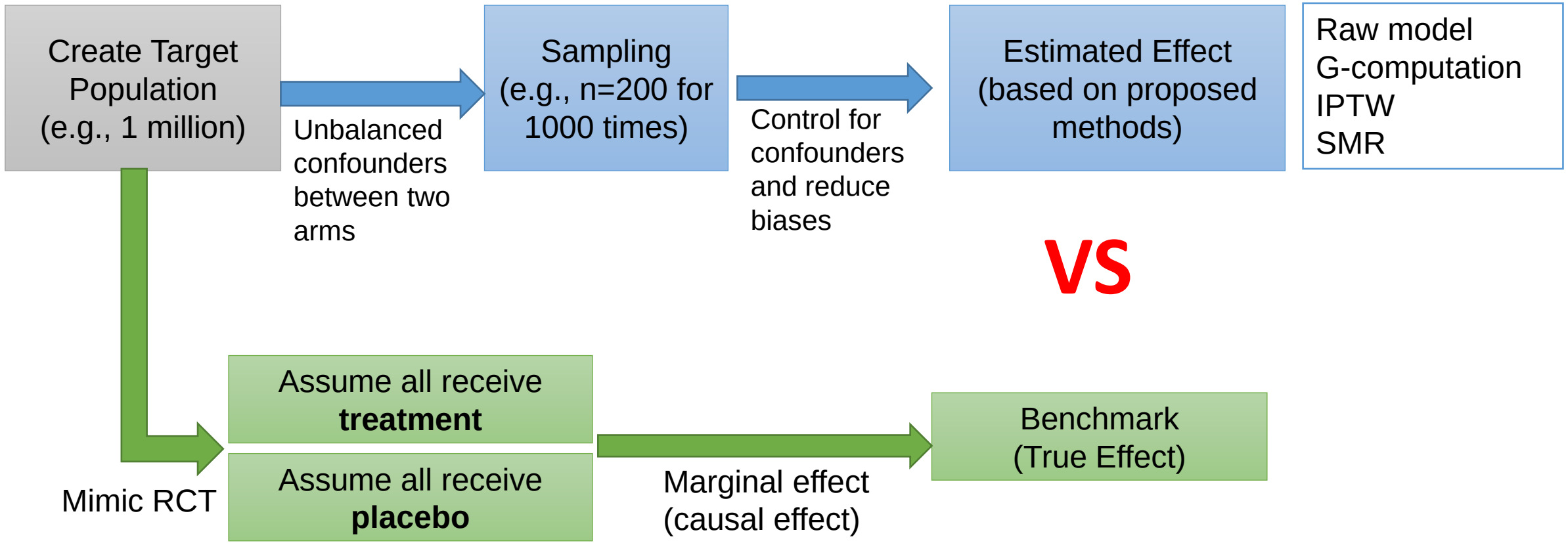


Objective: To evaluate two types of methods, propensity score (PS) based methods and g-computation in estimating treatment effects in a rare disease setting with small sample size using simulation

Methods

- Published data were used to set up treatment effect of potential intervention and confounders.
- Simulation was used to compare different techniques including unadjusted logistic regression, PS-based methods (inverse probability of treatment weighting (IPTW), standardized mortality/morbidity ratio weighting (SMR)), and g-computation to estimate the "**true**" treatment effect on overall survival (OS) in a specific target population using an external comparator design.
- Simulations of different scenarios were designed varying parameters such as effect size, total sample size and sample size ratio between arms, imbalance in confounders, and correlation between confounders.
- Accuracy and precision of estimates was evaluated by a combination of 95% confidence interval (CI) coverage, power, bias, and mean square error (MSE).

Framework For Simulation



Detailed Process of Setting up Benchmark Effect

*Use logistic regression as an example

1

Create a target population

Treatment, T (A=1, B=0); Confounders, X1-X3

Outcome, Y

Use logistic regression to generate the outcome variable Y

$$\text{logit}(p_y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_t * T$$

2

Target population

Confounders, X1-X3

Treatment A (T=1)

Treatment B (T=0)

Outcome, Y1 for treatment A
Outcome, Y0 for treatment B

Assume everyone in the population receives Treatment A, and apply the same coefficients to generate the outcome variable Y1

$$\text{logit}(p_{y1}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_t * 1$$

Assume everyone in the population receives Treatment B, and apply the estimated coefficients to generate the outcome variable Y0

$$\text{logit}(p_{y0}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_t * 0$$

3

Treatment A (T=1) **VS** Treatment B (T=0)

Estimate the effect of treatment (causal effect, marginal effect)

Risk Difference (RD) = p_{y1} - p_{y0}

Relative Risk (RR) = p_{y1} / p_{y0}

Odds Ratio (OR) = $[p_{y1}/(1-p_{y1})]$ / $[p_{y0}/(1-p_{y0})]$

Base Case Simulation Scenario

Parameters	Values
Effect size (hazard ratio HR)	0.5
Total sample size	150, 1000
Ratio (N_T vs. N_c)	1:4
Strength of confounders (association with the outcome)	X_1 weak, X_2 medium, X_3 strong
Unbalance of confounders (association with treatment arm)	Cor= 0.5 (medium)
Correlation between confounders	r= 0.2

In Small Sample Scenario, All 3 Methods Demonstrated Sufficient Control for Confounding With G-computation Performing Better

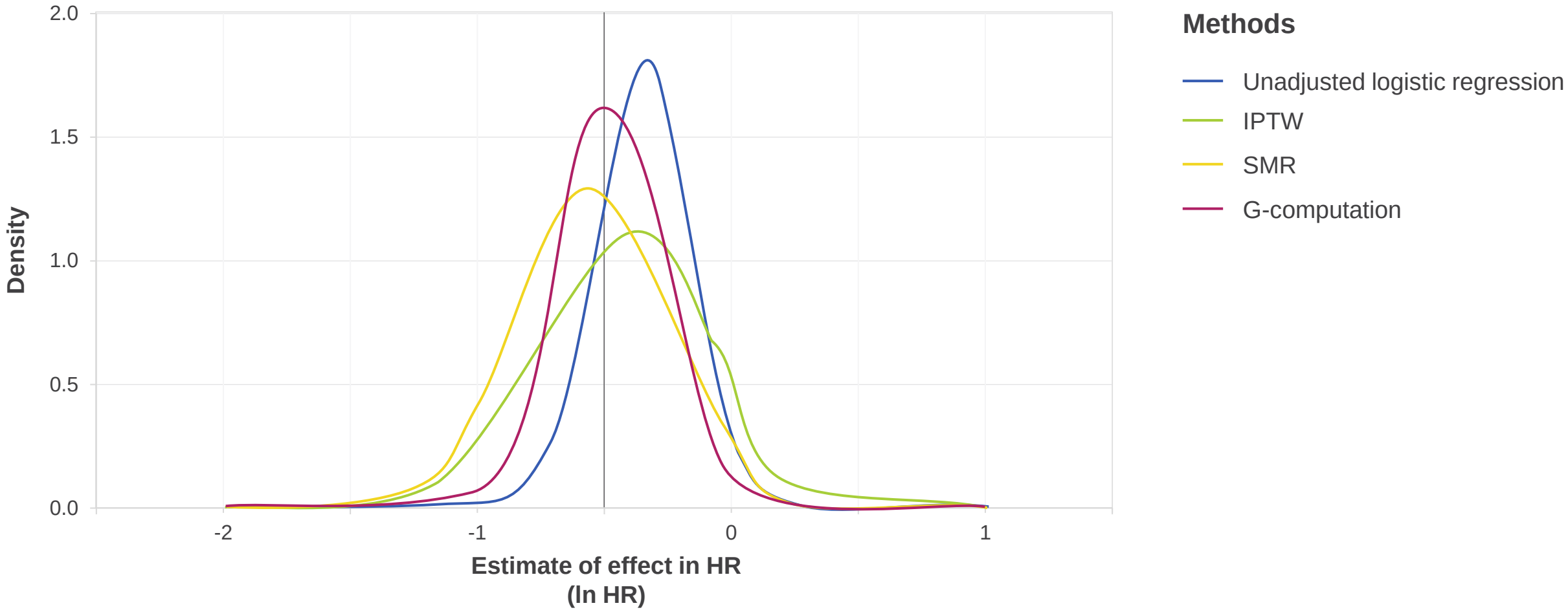
N_T vs. N_C: 30 vs 120; true HR=0.57

Method	95% CI Coverage, %	Power, %	Log (HR)		
			95% CI width	Bias	MSE
Unadjusted logistic regression	87.8	41.6	0.89	0.17	0.08
IPTW	65.6	62.6	0.73	0.07	0.17
SMR	72.8	75.9	0.73	-0.03	0.11
G-computation	93.5	69.3	0.90	-0.00	0.06

CI, confidence interval; HR, hazard ratio; MSE, mean squared error.

Distribution of OS Point Estimates in a Small Sample Scenario

N_T vs. N_C: 30 vs 120; true HR=0.57



In Large Sample Scenario, Among All the Methods G-computation Performed Better

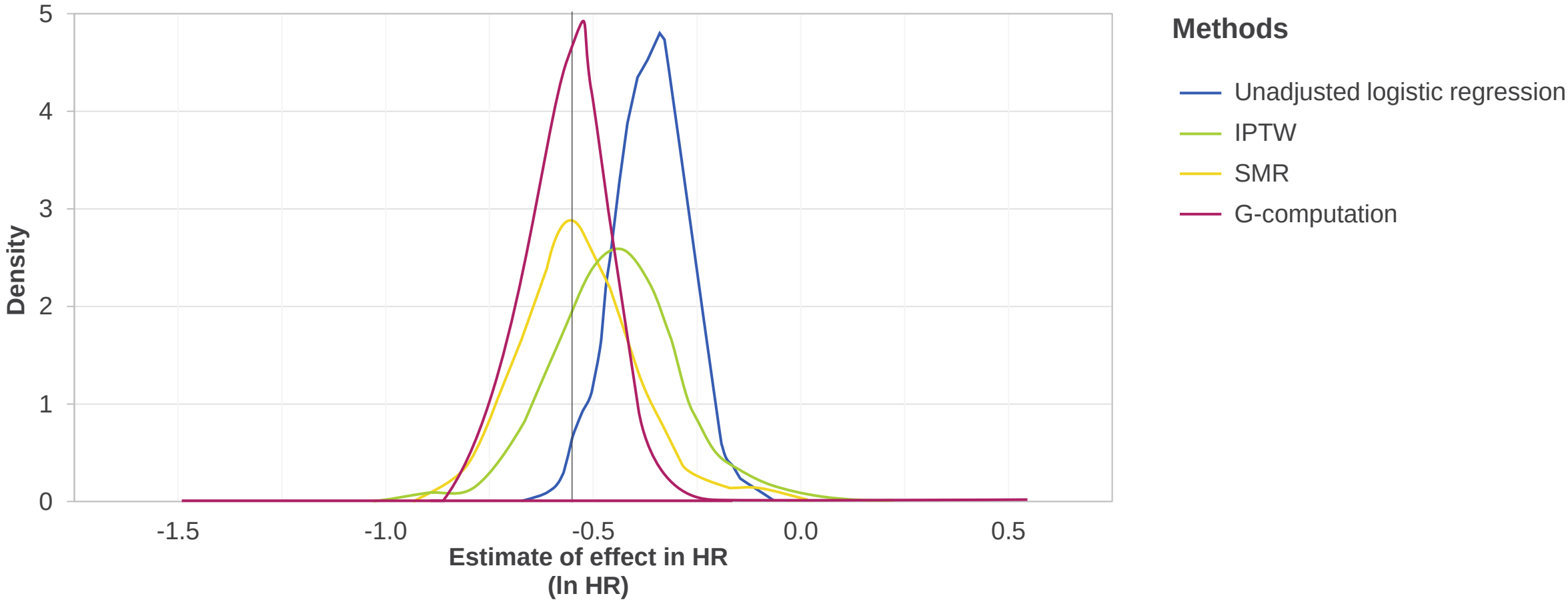
N_T vs. N_C: 200 vs 800; true HR=0.57

Method	95% CI Coverage, %	Power, %	Log (HR)		
			95% CI width	Bias	MSE
Unadjusted logistic regression	45.9	99.7	0.33	0.17	0.04
IPTW	52.5	97.4	0.27	0.10	0.04
SMR	65.9	98.9	0.27	0.00	0.02
G-computation	91.9	100.0	0.33	-0.01	0.01

CI, confidence interval; HR, hazard ratio; MSE, mean squared error.

Distribution of OS Point Estimates in a Large Sample Scenario

N_T vs. N_C: 200 vs 800; true HR=0.57



Conclusions

- In our analysis, two categories of methods demonstrated sufficient control for confounding and can be used for estimation of efficacy in small single arm trial when RCT is infeasible or unethical in rare diseases.
- PS-based methods (e.g., SMR & IPTW) have been shown to be suitable in the creation of the comparator arm, and one of the most commonly-used ones based on the number of publications per year.
- Newer methods such as g-computation may produce more efficient estimate; however, utility of this method in prospective evaluation of the efficacy should be considered carefully since it involves inclusion of the outcome.

Contact Information

Arie Barlev, PharmD, MS
Vice President, Global Value Access, and Outcomes
Atara Biotherapeutics, Inc.
abarlev@atarabio.com