Supporting Information for "Power and Sample Size for Observational Studies of Point Exposure Effects"

by

Bonnie E. Shook-Sa and Michael G. Hudgens

Web Appendix A: Proofs of Propositions

A.1 *Proof of Proposition 1*

The proof of Proposition 1 relies on standard estimating equation theory (see Stefanski and Boos, 2002). Let $O_i = (L_i, A_i, Y_i)$, $\mu = (\mu_1, \mu_0)$, and note $\hat{\mu} = (\hat{\mu}_1, \hat{\mu}_0)$ solves the estimating equations

$$\sum_{i=1}^{n} \Psi(O_i, \mu) = \left\{ \begin{array}{c} \sum_{i=1}^{n} \Psi_1(O_i, \mu) \\ \sum_{i=1}^{n} \Psi_0(O_i, \mu) \end{array} \right\} = \left\{ \begin{array}{c} \sum_{i=1}^{n} W_{1i}(Y_i - \mu_1)A_i \\ \sum_{i=1}^{n} W_{0i}(Y_i - \mu_0)(1 - A_i) \end{array} \right\} = 0$$

Let $\dot{\Psi}(O_i, \mu) = \partial\Psi(O_i, \mu)/\partial\mu$, $A(\mu) = E(-\dot{\Psi})$, and $B(\mu) = E\{\Psi(O_i, \mu)\Psi(O_i, \mu)^T\}$. It is straightforward to show $A(\mu)$ equals the 2×2 identify matrix and

$$B(\mu) = E \left\{ \begin{array}{cc} W_{1i}(Y_{1i} - \mu_1)^2 & 0 \\ 0 & W_{0i}(Y_{0i} - \mu_0)^2 \end{array} \right\}$$

Therefore as $n \to \infty$,

$$\sqrt{n} \left\{ \begin{pmatrix} \hat{\mu}_1 \\ \hat{\mu}_0 \end{pmatrix} - \begin{pmatrix} \mu_1 \\ \mu_0 \end{pmatrix} \right\} \xrightarrow{d} N\left(0, V(\mu)\right) \tag{A.1}$$

where $V(\mu) = A(\mu)^{-1}B(\mu)\{A(\mu)^{-1}\}^T = B(\mu)$. Without loss of generality, consider $a = 1$. Then $\sqrt{n}\,(\hat{\mu}_1 - \mu_1) \xrightarrow{d} N\left(0, \Sigma_1\right)$ where $\Sigma_1 = E\{W_{1i}(Y_{1i} - \mu_1)^2\}$. Dropping subscripts i for notational ease, note

$$\Sigma_1 = E(W_1 Y_1^2) + \mu_1^2 E(W_1) - 2\mu_1 E(W_1 Y_1) \tag{A.2}$$

As in Kong (1992), next define $R = E[\{W_1 - E(W_1)\}(Y_1 - \mu_1)^2]$ such that

$$R = E(W_1 Y_1^2) - 2\mu_1 E(W_1 Y_1) - E(Y_1^2)E(W_1) + 2\mu_1^2 E(W_1) \tag{A.3}$$

From (A.2) and (A.3) it follows that

$$\begin{aligned}\Sigma_1 &= E(W_1)E(Y_1^2) - E(W_1)\mu_1^2 + R = E(W_1)\sigma_1^2 + R \\ &= \sigma_1^2 E(W_1^2 A) + R = \sigma_1^2\left[\frac{E(W_1^2 A)}{E(W_1 A)}\right] + R \\ &= \sigma_1^2\left[\frac{E(W^2 A)}{\{E(WA)\}^2}\right] + R\end{aligned}$$

Bounds for R follow from the Cauchy-Schwarz inequality:

$$|R| = |Cov(W_1, Y_1^2 - 2\mu_1 Y_1)| \leqslant \sqrt{Var(W_1)Var(Y_1^2 - 2\mu_1 Y_1)}$$

A.2 *Proof of Proposition 2*

Let $1-\beta$ denote the power to detect a difference in causal means of size δ, i.e.,

$$\begin{aligned}1-\beta &= P(|t| > z_{1-\alpha/2} \mid ACE = \delta) \\ &= P\left(\frac{\widehat{ACE}-\delta}{\sqrt{Var_n(\widehat{ACE})}} > z_{1-\alpha/2} - \frac{\delta}{\sqrt{Var_n(\widehat{ACE})}}\middle| ACE = \delta\right) \\ &\quad + P\left(\frac{\widehat{ACE}-\delta}{\sqrt{Var_n(\widehat{ACE})}} < z_{\alpha/2} - \frac{\delta}{\sqrt{Var_n(\widehat{ACE})}}\middle| ACE = \delta\right)\end{aligned}$$

In large samples, $(\widehat{ACE} - ACE)\, Var_n(\widehat{ACE})^{-1/2}$ is approximately standard normal. Thus,

$$1-\beta \approx 1 - \Phi\left(z_{1-\alpha/2} - \frac{\delta}{\sqrt{Var_n(\widehat{ACE})}}\right) + \Phi\left(z_{\alpha/2} - \frac{\delta}{\sqrt{Var_n(\widehat{ACE})}}\right) \tag{A.4}$$

where $\Phi(*)$ represents the cumulative distribution function for the standard normal evaluated at $*$. Without loss of generality, assume $\delta > 0$. Then the second component on the right side of (A.4) will be less than $\alpha/2$ and often close to zero. Therefore,

$$z_\beta \approx z_{1-\alpha/2} - \frac{\delta}{\sqrt{Var_n(\widehat{ACE})}} \tag{A.5}$$

Define $k = P(A=1)/P(A=0)$. Given that $Var_n(\widehat{ACE}) = \{nP(A=1)\}^{-1}\sigma_{1,adj}^2 + \{nP(A=0)\}^{-1}\sigma_{0,adj}^2$ and solving (A.5) for n yields (7).

Web Appendix B: Additional Simulations

B.1 *Pilot Study Simulations*

As discussed in Section 4.4, simulations were conducted to evaluate the performance of the design effect approximation when pilot study data are available for Scenarios 1-4. Pilot samples of size $n_p = 100$ were selected from the superpopulation and then k and $\sigma^2_{a,adj}$ for $a \in \{0, 1\}$ were estimated from the pilot sample. Based on these estimates and δ from Table 1, n_{deff} was calculated and Steps (ii) - (vi) from Section 4.3 were followed based on this sample size.

Web Table 1 presents the distribution of the estimated variances ($\hat{\sigma}^2_a$), design effects ($\widehat{deff}^a_w$) and adjusted variances ($\hat{\sigma}^2_{a,adj}$) for $a \in \{0, 1\}$ along with the sample size n_{deff} across the $R = 2000$ simulated samples for Scenarios 1b-4b. With pilot samples of size $n_p = 100$, there is considerable variation in the estimated design effects ($\widehat{deff}^0_w$ and $\widehat{deff}^1_w$) and sample size n_{deff} calculated using estimates from the pilot study. As anticipated, median estimated design effects and the median sample sizes are similar to those presented in Table 2. The results of the simulation study are presented in Web Figure 1. These results are similar to those presented in Figure 1A, when no pilot data were available.

B.2 *Null Simulations*

For each scenario discussed in the main text, simulations were conducted under the null hypothesis to estimate the Type I error rate empirically. That is, simulations were conducted as described in Section 4, with n_{deff} taken from Table 2, but $ACE = 0$ in the superpopulation. Web Figure 2 contains the results of the null simulations. For all scenarios, the empirical Type I error rate was approximately 5%.

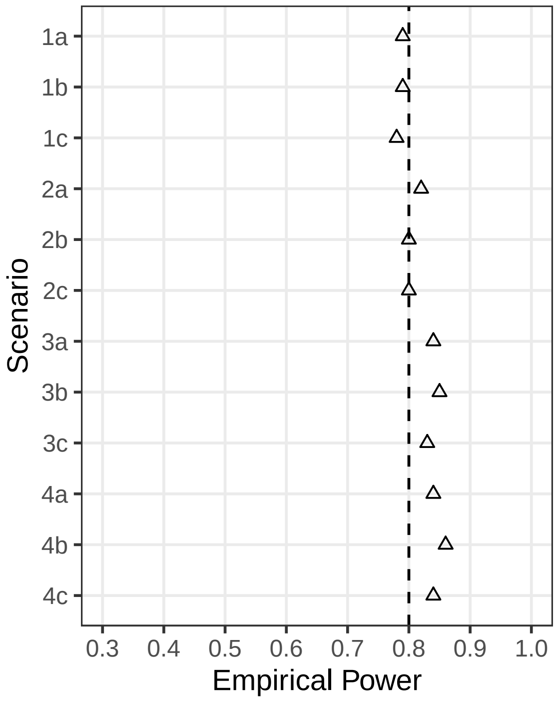
B.3 *Simulations for Section 6*

Simulations were conducted using $\widehat{deff}^{a}_{w,rem}$ to calculate required sample sizes for Scenarios 1-4. That is, sample sizes were computed using the methods discussed in Section 6 rather than ignoring the remainder term Er_a. For each iteration of the simulation, a pilot sample of size $n_p \in \{100, 500, 1000\}$ was selected from the superpopulation and k and $\sigma^2_{a,adj}$ for $a \in \{0, 1\}$ were estimated from the pilot sample. Adjusted variances $\sigma^2_{a,adj}$ were estimated by $\hat{\sigma}^2_{a,adj} = \hat{\sigma}^2_a \widehat{deff}^{a}_{w,rem}$. Based on these estimates and δ from Table 1, n_{deff} was calculated and Steps (ii) - (vi) from Section 4.3 were followed based on this sample size. The results of these simulations are shown in Web Figure 3 and Web Figure 4.

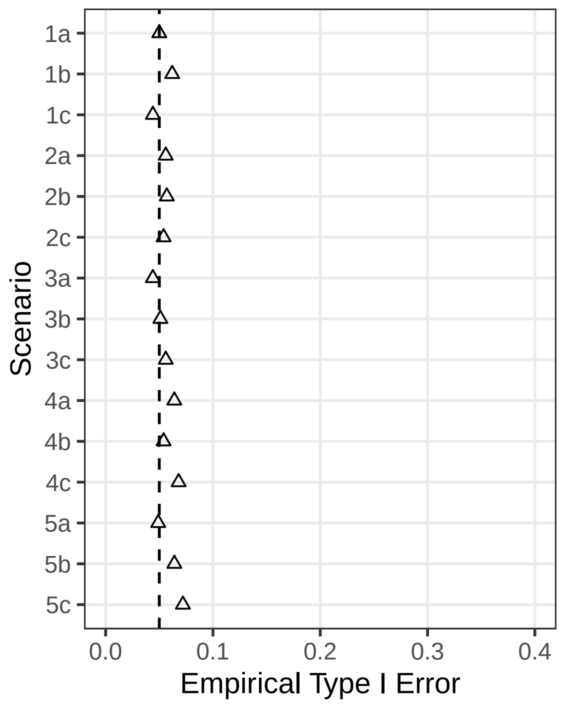
References

Kong, A. (1992). A note on importance sampling using standardized weights. *University of Chicago, Dept. of Statistics, Tech. Rep* **348,** 1–4.

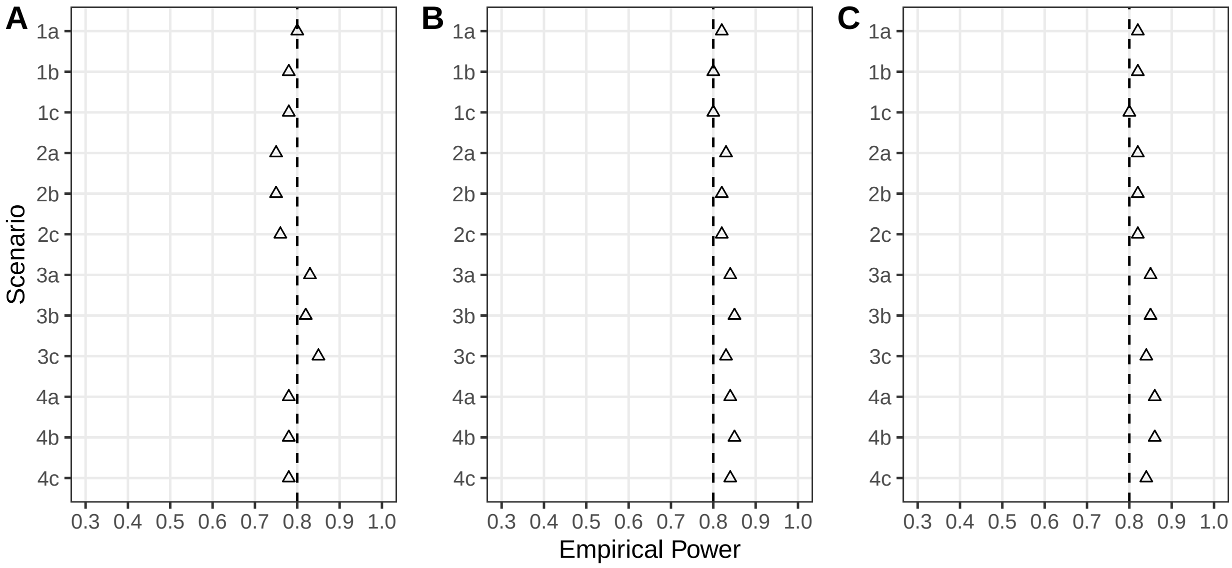
Stefanski, L. A. and Boos, D. D. (2002). The calculus of M-estimation. *The American Statistician* **56,** 29–38.



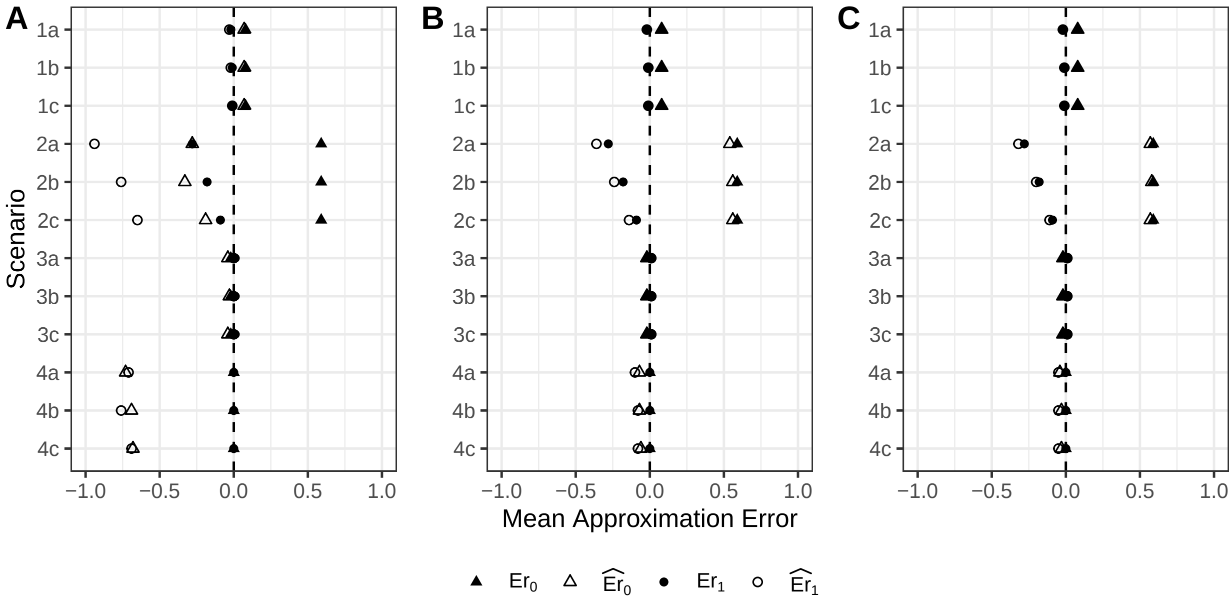
Web Figure 1. Empirical power from the simulation study described in Web Appendix B.1 by scenario across $R = 2000$ samples. Empirical power is the proportion of simulated samples in which the p-value for testing $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ was less than $\alpha = 0.05$ for the MSM $E(Y_a) = \beta_0 + \beta_1 a$, based on the sample size estimated using pilot study data.



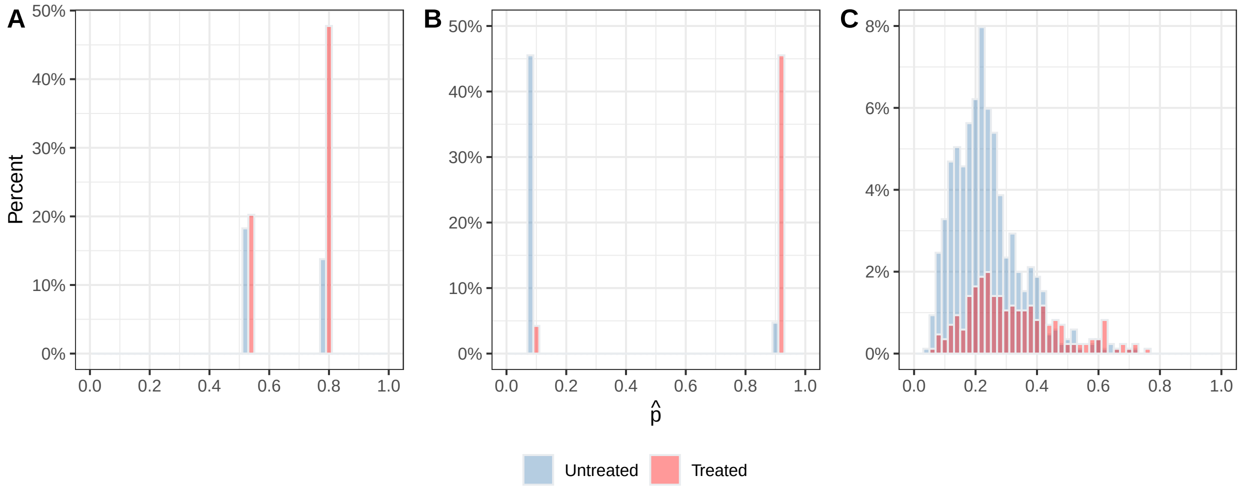
Web Figure 2. Empirical Type I error from the simulation study described in Web Appendix B.2 by scenario across $R = 2000$ samples. Empirical Type I Error is the proportion of simulated samples in which the p-value for testing $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ was less than $\alpha = 0.05$ for the MSM $E(Y_a) = \beta_0 + \beta_1 a$, based on sample size n_{deff} from Table 2 when $ACE = 0$. (Scenario 5c excludes 2 simulations in which the geex package failed to converge when estimating the standard error of $\widehat{ACE}$.)



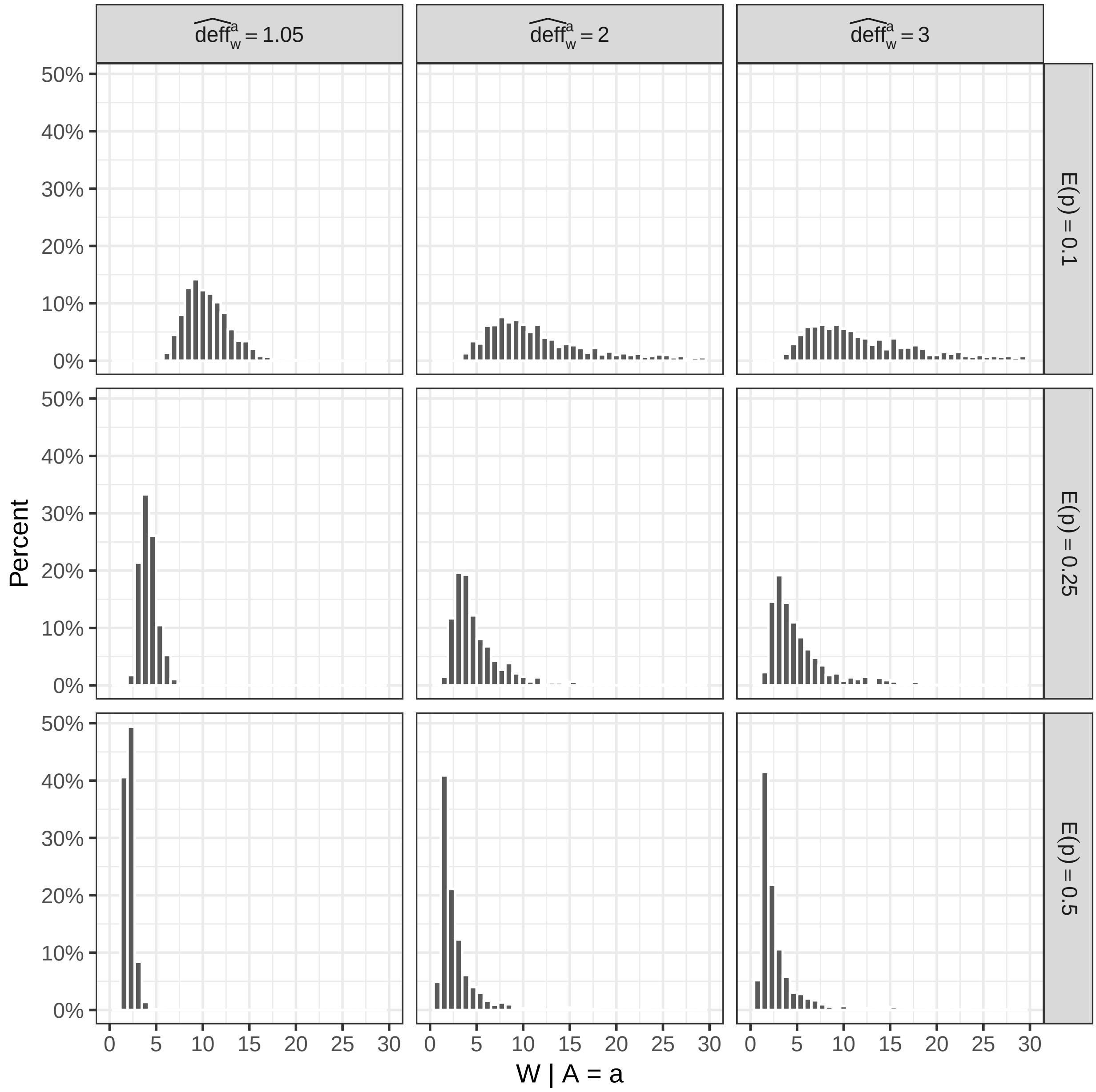
Web Figure 3. Empirical power from the simulation study described in Web Appendix B.3 by scenario across $R = 2000$ samples for (A) $n_p = 100$, (B) $n_p = 500$, and (C) $n_p = 1000$. Empirical power is the proportion of simulated samples in which the p-value for testing $H_0 : \beta_1 = 0$ versus $H_1 : \beta_1 \neq 0$ was less than $\alpha = 0.05$ for the MSM $E(Y_a) = \beta_0 + \beta_1 a$, based on the sample size calculated using estimates from the pilot study. (Scenarios 2a-2c and 4a-4c with $n_p = 100$ excluded between 13 and 30 simulations in which the design effect was estimated to be negative.)



Web Figure 4. Mean approximation error estimated from the simulation study described in Web Appendix B.3 by scenario across $R = 2000$ samples for (A) $n_p = 100$, (B) $n_p = 500$, and (C) $n_p = 1000$. Approximation errors were estimated from each pilot sample and were averaged across the $R = 2000$ simulated samples. (Scenarios 2a-2c and 4a-4c with $n_p = 100$ excluded between 13 and 30 simulations in which the design effect was estimated to be negative.)



Web Figure 5. Estimated propensity scores from a single simulation by scenario for (A) Scenarios 1 and 3, (B) Scenarios 2 and 4, and (C) Scenario 5. Note Scenarios 1 and 3 have the same distribution of propensity scores because the joint distribution of A and L is the same. Likewise, Scenarios 2 and 4 have the same distribution of propensity scores. Jitter is applied to the estimated propensity scores in the treatment group for Scenarios 1 and 2.



Web Figure 6. Examples of weight distributions for various approximated design effects and mean propensity score $E(p)$. Distributions were generated by taking reciprocals of $N_a = 1000$ random draws from beta distributions with shape parameters set to achieve the desired $E(p)$ and design effect.

Web Table 1

Distribution of estimated variances, design effects, adjusted variances, and required sample sizes for select pilot study simulation scenarios across $R = 2000$ samples.

Scenario		$\hat{\sigma}_0^2$	$\hat{\sigma}_1^2$	$\widehat{deff}_w^0$	$\widehat{deff}_w^1$	$\hat{\sigma}_{0,adj}^2$	$\hat{\sigma}_{1,adj}^2$	n_{deff}
	min	0.0241	0.1647	1.00	1.00	0.0267	0.1727	157
	p25	0.1667	0.2349	1.06	1.02	0.1857	0.2437	319
1b	p50	0.1966	0.2431	1.12	1.04	0.2203	0.2519	354
	p75	0.2188	0.2482	1.20	1.07	0.2499	0.2598	392
	max	0.2500	0.2500	2.72	1.44	0.6157	0.3509	796
	min	0.0109	0.1245	1.00	1.00	0.0247	0.2304	326
	p25	0.1432	0.2110	2.27	2.30	0.3555	0.5180	672
2b	p50	0.1853	0.2349	2.78	2.83	0.4863	0.6448	834
	p75	0.2201	0.2463	3.66	3.70	0.6629	0.8278	1071
	max	0.2500	0.2500	13.15	13.37	3.1739	2.7077	3410
	min	58.1	136.9	1.00	1.00	66.7	140.4	183
	p25	132.5	241.5	1.07	1.02	151.7	251.3	271
3b	p50	159.5	272.3	1.12	1.04	179.7	286.1	306
	p75	189.5	308.7	1.19	1.07	214.8	323.9	343
	max	334.8	445.1	2.10	1.39	463.6	547.5	647
	min	49.8	69.6	1.00	1.00	96.1	173.4	280
	p25	119.3	200.3	2.27	2.28	325.7	529.7	613
4b	p50	151.2	250.8	2.81	2.79	432.2	723.5	779
	p75	191.6	311.8	3.69	3.70	601.4	1010.0	1028
	max	493.9	869.1	13.48	13.97	4509.9	7779.1	5453