Web-based Supplementary Materials for
"Assessing effects of cholera vaccination
in the presence of interference"

by

Carolina Perez-Heydrich, Michael G. Hudgens,* M. Elizabeth Halloran,
John D. Clemens, Mohammad Ali, and Michael E. Emch

* email: mhudgens@bios.unc.edu

Web Appendix A

Here the asymptotic distributions of the IPW estimators in the main paper are derived using M-estimation theory as presented in Stefanski and Boos (2002). In Section A.1 the propensity scores are assumed known. This assumption is relaxed in Section A.2. Some additional comments regarding variance estimation are given in Section A.3. All equation numbers refer to equations within this appendix and do not refer to equations in the main text. Throughout it is assumed the m groups constitute a random sample such that the observable random variables $(\mathbf{Y}_i, \mathbf{A}_i, \mathbf{B}_i, \mathbf{X}_i)$ for $i = 1, \dots, m$ are independent and identically distributed.

A.1 Known Propensity Score

First we derive the asymptotic distribution of $\widehat{Y}^{ipw}(a; \alpha)$ when the propensity scores are known. To more closely mimic the notation in Stefanski and Boos, let $\hat{\theta}_{a,\alpha} = \widehat{Y}^{ipw}(a; \alpha)$ as defined in the main text. That is,

$$\hat{\theta}_{a,\alpha} = m^{-1} \sum_i \widehat{Y}_i^{ipw}(a; \alpha)$$

where here and in the sequel $\sum_i = \sum_{i=1}^m$ and

$$\widehat{Y}_i^{ipw}(a; \alpha) = \frac{\sum_{j=1}^{n_i} \pi(\mathbf{A}_{i,-j}; \alpha) I(A_{ij} = a) Y_{ij}}{\Pr(\mathbf{A}_i | \mathbf{X}_i; \boldsymbol{\theta}_x, \theta_s) n_i}.$$

In contrast to the main text, here we denote the parameters of the propensity score model by $\boldsymbol{\theta}_x, \theta_s$ instead of $\boldsymbol{\phi}_a, \phi_b$ to better match the notation in

Stefanski and Boos. That is, $\boldsymbol{\theta}_x$ is a $(p \times 1)$ vector of regression coefficients and θ_s is the variance of the random effect b_i. For now we are assuming $\boldsymbol{\theta}_x, \theta_s$ are known; in the next section below we will consider the setting where these parameters are unknown and must be estimated from the data. Equivalently we can express the IPW estimator as

$$\hat{\theta}_{a,\alpha} = m^{-1} \sum_i \frac{g(\mathbf{Y}_i, \mathbf{A}_i, a, \alpha)}{\Pr(\mathbf{A}_i | \mathbf{X}_i; \boldsymbol{\theta}_x, \theta_s)} \tag{1}$$

where

$$g(\mathbf{Y}_i, \mathbf{A}_i, a, \alpha) = \sum_{j=1}^{n_i} \pi(\mathbf{A}_{i,-j}; \alpha) I(A_{ij} = a) Y_{ij}/n_i.$$

Note the estimator $\hat{\theta}_{a,\alpha}$ is a solution (for θ) to the estimating equation

$$\sum_i \psi_{a,\alpha}(\mathbf{Y}_i, \mathbf{A}_i, \mathbf{X}_i, \theta) = 0$$

where

$$\psi_{a,\alpha}(\mathbf{Y}_i, \mathbf{A}_i, \mathbf{X}_i, \theta) = \frac{g(\mathbf{Y}_i, \mathbf{A}_i, a, \alpha)}{\Pr(\mathbf{A}_i | \mathbf{X}_i; \boldsymbol{\theta}_x, \theta_s)} - \theta.$$

Therefore, by M-estimation theory, $\hat{\theta}_{a,\alpha} \xrightarrow{p} \theta_{a,\alpha}$ and $\sqrt{m}(\hat{\theta}_{a,\alpha} - \theta_{a,\alpha})$ converges in distribution to $N(0, \Sigma)$ where the variance matrix Σ has the sandwich form

$$\Sigma = U(\theta_{a,\alpha})^{-1} V(\theta_{a,\alpha}) U(\theta_{a,\alpha})^{-1}$$

where

$$U(\theta_{a,\alpha}) = E[-\dot{\psi}_{a,\alpha}(\mathbf{Y}_i, \mathbf{A}_i, \mathbf{X}_i, \theta_{a,\alpha})]$$
$$V(\theta_{a,\alpha}) = E[\psi_{a,\alpha}(\mathbf{Y}_i, \mathbf{A}_i, \mathbf{X}_i, \theta_{a,\alpha})^2]$$
$$\dot{\psi}_{a,\alpha}(\mathbf{Y}_i, \mathbf{A}_i, \mathbf{X}_i, \theta) = \partial \psi_{a,\alpha}(\mathbf{Y}_i, \mathbf{A}_i, \mathbf{X}_i, \theta)/\partial \theta$$

and $\theta_{a,\alpha}$ is the true parameter value defined by

$$\int \psi_{a,\alpha}(\mathbf{y}, \mathbf{a}, \mathbf{x}, \theta_{a,\alpha}) dF(\mathbf{y}, \mathbf{a}, \mathbf{x}) = 0,$$

i.e.,

$$\theta_{a,\alpha} = \int \frac{g(\mathbf{y}, \mathbf{a}, a, \alpha)}{\Pr(\mathbf{a} | \mathbf{x}; \boldsymbol{\theta}_x, \theta_s)} dF(\mathbf{y}, \mathbf{a}, \mathbf{x}),$$

where F denotes the cumulative distribution function of $\mathbf{Y}_i, \mathbf{A}_i, \mathbf{X}_i$. In this simple case $\dot{\psi}_{a,\alpha}(\mathbf{Y}_i, \mathbf{A}_i, \mathbf{X}_i, \theta) = -1$, implying $U(\theta_{a,\alpha})$=1 and therefore $\Sigma = V(\theta_{a,\alpha})$ which can be consistently estimated by the empirical estimator

$$V_m(\mathbf{Y}, \mathbf{A}, \mathbf{X}) = \frac{1}{m} \sum_i \psi_{a,\alpha}(\mathbf{Y}_i, \mathbf{A}_i, \mathbf{X}_i, \hat{\theta}_{a,\alpha})^2.$$

This implies, using the notation in the main paper, that the large sample variance of $\widehat{Y}^{ipw}(a;\alpha)$ can be estimated by

$$\frac{1}{m^2} \sum_i \left\{ \widehat{Y}_i^{ipw}(a;\alpha) - \widehat{Y}^{ipw}(a;\alpha) \right\}^2.$$

Note the form of this variance estimator intuitively makes sense because the estimator (1) is essentially just a sample mean across groups.

Large sample variance estimators of the direct, indirect, total, and overall IPW estimators can be derived analogously. In particular, note that the direct effect estimator $\widehat{DE}(\alpha)$ is a solution to the estimating equation

$$\sum_i \psi_{de;\alpha}(\mathbf{Y}_i, \mathbf{A}_i, \mathbf{X}_i, \theta) = 0$$

where

$$\psi_{de;\alpha}(\mathbf{Y}_i, \mathbf{A}_i, \mathbf{X}_i, \theta) = \frac{g(\mathbf{Y}_i, \mathbf{A}_i, 0, \alpha) - g(\mathbf{Y}_i, \mathbf{A}_i, 1, \alpha)}{\Pr(\mathbf{A}_i | \mathbf{X}_i; \boldsymbol{\theta}_x, \theta_s)} - \theta,$$

implying the large sample variance of $\widehat{DE}(\alpha)$ can be estimated by

$$\frac{1}{m^2} \sum_i \left\{ \widehat{DE}_i(\alpha) - \widehat{DE}(\alpha) \right\}^2, \tag{2}$$

i.e., by the between group sample variance of the direct effect estimates. Likewise the indirect, total, and overall effect IPW estimators are solutions to the estimating equations

$$\psi_{ie;\alpha,\alpha'}(\mathbf{Y}_i, \mathbf{A}_i, \mathbf{X}_i, \theta) = \frac{g(\mathbf{Y}_i, \mathbf{A}_i, 0, \alpha) - g(\mathbf{Y}_i, \mathbf{A}_i, 0, \alpha')}{\Pr(\mathbf{A}_i | \mathbf{X}_i; \boldsymbol{\theta}_x, \theta_s)} - \theta,$$

$$\psi_{te;\alpha,\alpha'}(\mathbf{Y}_i, \mathbf{A}_i, \mathbf{X}_i, \theta) = \frac{g(\mathbf{Y}_i, \mathbf{A}_i, 0, \alpha) - g(\mathbf{Y}_i, \mathbf{A}_i, 1, \alpha')}{\Pr(\mathbf{A}_i | \mathbf{X}_i; \boldsymbol{\theta}_x, \theta_s)} - \theta,$$

$$\psi_{oe;\alpha,\alpha'}(\mathbf{Y}_i, \mathbf{A}_i, \mathbf{X}_i, \theta) = \frac{\tilde{g}(\mathbf{Y}_i, \mathbf{A}_i, \alpha) - \tilde{g}(\mathbf{Y}_i, \mathbf{A}_i, \alpha')}{\Pr(\mathbf{A}_i|\mathbf{X}_i; \boldsymbol{\theta}_x, \theta_s)} - \theta,$$

where

$$\tilde{g}(\mathbf{Y}_i, \mathbf{A}_i, \alpha) = \sum_{j=1}^{n_i} \pi(\mathbf{A}_i; \alpha) Y_{ij}/n_i.$$

It follows that estimators of the same form as (2) can be used to estimate the asymptotic variance of the indirect, total, and overall effect estimators.

A.2 Unknown Propensity Scores

As noted in the main text, in observational settings the true propensity score will never be known and must be estimated. In this case, a vector of estimating equations can be used to derive large sample variance estimators. We first consider estimating the asymptotic variance of $\hat{\theta}_{a,\alpha} = \hat{Y}^{ipw}(a;\alpha)$. To begin, write the log likelihood for the mixed effects logit model of the probability of participating in the study as $\sum_i l(\mathbf{B}_i, \mathbf{X}_i, \boldsymbol{\theta}_x, \theta_s)$ where

$$l(\mathbf{B}_i, \mathbf{X}_i, \boldsymbol{\theta}_x, \theta_s) = \log\left[\int \prod_{j=1}^{n_i} h_{ij}(b_i; \boldsymbol{\theta}_x)^{B_{ij}} \{1 - h_{ij}(b_i; \boldsymbol{\theta}_x)\}^{1-B_{ij}} f_b(b_i; \theta_s) db_i\right]$$

and $h_{ij}(b_i; \boldsymbol{\theta}_x) = \Pr(B_{ij} = 1|\mathbf{X}_{ij}, b_i) = \text{logit}^{-1}(\mathbf{X}_{ij}\boldsymbol{\theta}_x + b_i)$. Estimates that maximize this likelihood can be obtained using software for generalized linear mixed effects models, such as the `glmer` function in the R package `lme4` (Bates et al. 2011). Estimates $\hat{\boldsymbol{\theta}}_x, \hat{\theta}_s$ that maximize the log likelihood are solutions to the score equations

$$\sum_i \psi_{xk}(\mathbf{B}_i, \mathbf{X}_i, \boldsymbol{\theta}_x, \theta_s) = 0 \text{ for } k = 1, \ldots, p$$

$$\sum_i \psi_s(\mathbf{B}_i, \mathbf{X}_i, \boldsymbol{\theta}_x, \theta_s) = 0$$

where $\psi_{xk}(\mathbf{B}_i, \mathbf{X}_i, \boldsymbol{\theta}_x, \theta_s) = \partial l(\mathbf{B}_i, \mathbf{X}_i, \boldsymbol{\theta}_x, \theta_s)/\partial\theta_{xk}$, θ_{xk} denotes element k of $\boldsymbol{\theta}_x$, and $\psi_s(\mathbf{B}_i, \mathbf{X}_i, \boldsymbol{\theta}_x, \theta_s) = \partial l(\mathbf{B}_i, \mathbf{X}_i, \boldsymbol{\theta}_x, \theta_s)/\partial\theta_s$.

Therefore when the propensity scores are not assumed known but instead are estimated from the data, the corresponding estimator $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\theta}}_x, \hat{\theta}_s, \hat{\theta}_{a,\alpha})$ is a solution to the vector equation

$$\sum_i \boldsymbol{\psi}(\mathbf{O}_i, \boldsymbol{\theta}) = \mathbf{0}$$

where $\mathbf{O}_i = (\mathbf{Y}_i, \mathbf{A}_i, \mathbf{B}_i, \mathbf{X}_i)$ and

$$\boldsymbol{\psi}(\mathbf{O}_i, \boldsymbol{\theta}) = \begin{pmatrix} \psi_{x1}(\mathbf{B}_i, \mathbf{X}_i, \boldsymbol{\theta}_x, \theta_s) \\ \vdots \\ \psi_{xp}(\mathbf{B}_i, \mathbf{X}_i, \boldsymbol{\theta}_x, \theta_s) \\ \psi_s(\mathbf{B}_i, \mathbf{X}_i, \boldsymbol{\theta}_x, \theta_s) \\ \psi_{a,\alpha}(\mathbf{Y}_i, \mathbf{A}_i, \mathbf{X}_i, \theta_{a,\alpha}) \end{pmatrix} \tag{3}$$

By M-estimation theory it follows that $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$ and $\sqrt{m}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ converges in distribution to a multivariate normal distribution $N(\mathbf{0}, \boldsymbol{\Sigma})$ where

$$\boldsymbol{\Sigma} = \mathbf{U}(\boldsymbol{\theta}_0)^{-1}\mathbf{V}(\boldsymbol{\theta}_0)\{\mathbf{U}(\boldsymbol{\theta}_0)^{-1}\}^T$$
$$\mathbf{U}(\boldsymbol{\theta}_0) = E\{-\dot{\boldsymbol{\psi}}(\mathbf{O}_i, \boldsymbol{\theta}_0)\}$$
$$\mathbf{V}(\boldsymbol{\theta}_0) = E\{\boldsymbol{\psi}(\mathbf{O}_i, \boldsymbol{\theta}_0)\boldsymbol{\psi}(\mathbf{O}_i, \boldsymbol{\theta}_0)^T\}$$
$$\dot{\boldsymbol{\psi}}(\mathbf{O}_i, \boldsymbol{\theta}) = \partial\boldsymbol{\psi}(\mathbf{O}_i, \boldsymbol{\theta})/\partial\boldsymbol{\theta}^T$$

and the true parameter vector $\boldsymbol{\theta}_0$ is defined as the solution to the equation

$$\int \boldsymbol{\psi}(\mathbf{o}, \boldsymbol{\theta}_0)dF(\mathbf{o}) = \mathbf{0}$$

where here F denotes the cumulative distribution function of $\mathbf{O}_i$.

Replacing $\mathbf{U}(\boldsymbol{\theta}_0)$ and $\mathbf{V}(\boldsymbol{\theta}_0)$ with empirical estimators yields the empirical sandwich variance estimator

$$\boldsymbol{\Sigma}_m = \mathbf{U}_m(\hat{\boldsymbol{\theta}})^{-1}\mathbf{V}_m(\hat{\boldsymbol{\theta}})\{\mathbf{U}_m(\hat{\boldsymbol{\theta}})^{-1}\}^T$$

where

$$\mathbf{U}_m(\hat{\boldsymbol{\theta}}) = \frac{1}{m}\sum_i\{-\dot{\boldsymbol{\psi}}(\mathbf{O}_i, \hat{\boldsymbol{\theta}})\}$$

and

$$\mathbf{V}_m(\hat{\boldsymbol{\theta}}) = \frac{1}{m}\sum_i\left\{\boldsymbol{\psi}(\mathbf{O}_i, \hat{\boldsymbol{\theta}})\boldsymbol{\psi}(\mathbf{O}_i, \hat{\boldsymbol{\theta}})^T\right\}$$

The $(p+2, p+2)$ element (i.e., bottom right) of $\boldsymbol{\Sigma}_m$ multiplied by m^{-1} gives the estimated asymptotic variance of $\hat{\theta}_{a,\alpha} = \hat{Y}^{ipw}(a;\alpha)$. The estimated large sample variance for the direct, indirect, total, and overall effect estimators can be found in an analogous fashion by replacing the estimating equation $\psi_{a,\alpha}$ in (3) with the estimating equation corresponding to the effect of interest; for instance, $\psi_{de,\alpha}$ would be used in place of $\psi_{a,\alpha}$ in order to estimate the asymptotic variance of $\widehat{DE}(\alpha)$.

A.3 Variance Estimation

Note by definition $\dot{\boldsymbol{\psi}}(\mathbf{O}_i, \boldsymbol{\theta})$ equals the $(p+2) \times (p+2)$ matrix

$$\begin{bmatrix} \partial\psi_{x1}/\partial\theta_{x1} & \cdots & \partial\psi_{x1}/\partial\theta_{xp} & \partial\psi_{x1}/\partial\theta_{s} & \partial\psi_{x1}/\partial\theta_{a,\alpha} \\ \vdots & & \vdots & \vdots & \vdots \\ \partial\psi_{xp}/\partial\theta_{x1} & \cdots & \partial\psi_{xp}/\partial\theta_{xp} & \partial\psi_{xp}/\partial\theta_{s} & \partial\psi_{xp}/\partial\theta_{a,\alpha} \\ \partial\psi_{s}/\partial\theta_{x1} & \cdots & \partial\psi_{s}/\partial\theta_{xp} & \partial\psi_{s}/\partial\theta_{s} & \partial\psi_{s}/\partial\theta_{a,\alpha} \\ \partial\psi_{a,\alpha}/\partial\theta_{x1} & \cdots & \partial\psi_{a,\alpha}/\partial\theta_{xp} & \partial\psi_{a,\alpha}/\partial\theta_{s} & \partial\psi_{a,\alpha}/\partial\theta_{a,\alpha} \end{bmatrix} \tag{4}$$

where ψ_{xk} is shorthand for $\psi_{xk}(\mathbf{B}_i, \mathbf{X}_i, \boldsymbol{\theta}_x, \theta_s)$ and similarly for ψ_s and $\psi_{a,\alpha}$. The matrix (4) simplifies slightly to

$$\begin{bmatrix} \partial\psi_{x1}/\partial\theta_{x1} & \cdots & \partial\psi_{x1}/\partial\theta_{xp} & \partial\psi_{x1}/\partial\theta_{s} & 0 \\ \vdots & & \vdots & \vdots & \vdots \\ \partial\psi_{xp}/\partial\theta_{x1} & \cdots & \partial\psi_{xp}/\partial\theta_{xp} & \partial\psi_{xp}/\partial\theta_{s} & 0 \\ \partial\psi_{s}/\partial\theta_{x1} & \cdots & \partial\psi_{s}/\partial\theta_{xp} & \partial\psi_{s}/\partial\theta_{s} & 0 \\ \partial\psi_{a,\alpha}/\partial\theta_{x1} & \cdots & \partial\psi_{a,\alpha}/\partial\theta_{xp} & \partial\psi_{a,\alpha}/\partial\theta_{s} & -1 \end{bmatrix}$$

Therefore, using block matrix notation, we can write $\mathbf{U}(\boldsymbol{\theta}_0)$ as

$$\begin{bmatrix} \mathbf{U}_{11} & \mathbf{0} \\ \mathbf{U}_{21} & 1 \end{bmatrix}$$

where $\mathbf{U}_{11}$ is a $(p+1) \times (p+1)$ matrix and $\mathbf{U}_{21}$ is a $1 \times (p+1)$ vector. Similarly write $\mathbf{V}(\boldsymbol{\theta}_0)$ as

$$\begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{21}^T \\ \mathbf{V}_{21} & V_{22} \end{bmatrix}$$

where $\mathbf{V}_{11}$ is a $(p+1) \times (p+1)$ matrix, $\mathbf{V}_{21}$ is a $1 \times (p+1)$ vector, and V_{22} is a scalar. Then

$$\mathbf{U}(\boldsymbol{\theta}_0)^{-1} = \begin{bmatrix} \mathbf{U}_{11}^{-1} & \mathbf{0} \\ -\mathbf{U}_{21}\mathbf{U}_{11}^{-1} & 1 \end{bmatrix}$$

Because $\mathbf{U}_{11}$ and $\mathbf{V}_{11}$ correspond to the score equations of the log likelihood function of the mixed effects model, it follows that $\mathbf{U}_{11} = \mathbf{V}_{11}$ (Stefanski and Boos 2002). Therefore by straightforward linear algebra it follows that

$$\boldsymbol{\Sigma} = \begin{bmatrix} \mathbf{V}_{11}^{-1} & 0 \\ (-\mathbf{U}_{21} + \mathbf{V}_{21})\mathbf{V}_{11}^{-1} & (\mathbf{U}_{21} - 2\mathbf{V}_{21})\mathbf{V}_{11}^{-1}\mathbf{U}_{21}^T + V_{22} \end{bmatrix}$$

This suggests estimating the asymptotic variance of $\hat{\theta}_{a,\alpha}$ by

$$\frac{1}{m}\left\{(\hat{\mathbf{U}}_{21} - 2\hat{\mathbf{V}}_{21})\hat{\mathbf{V}}_{11}^{-1}\hat{\mathbf{U}}_{21}^{T} + \hat{V}_{22}\right\} \tag{5}$$

where $\hat{\mathbf{U}}_{21}$ denotes the first $p+1$ terms in the bottom row of $\mathbf{U}_m(\hat{\theta})$, and $\hat{\mathbf{V}}_{21}$, $\hat{\mathbf{V}}_{11}$, and $\hat{V}_{22}$ are the analogous submatrices (or elements) of $\mathbf{V}_m(\hat{\theta})$. Note (5) has the advantage of not requiring computation of the entire matrix $\mathbf{U}_m(\hat{\theta})$, for which the upper left $(p+1)\times(p+1)$ submatrix entails second derivatives of $l(\mathbf{B}_i, \mathbf{X}_i, \theta_x, \theta_s)$. For the results in the main text (i.e., the simulation study and cholera vaccine trial analysis), the derivatives required for computing (5) were evaluated numerically using the `grad` function in the numDeriv R package (Gilbert 2012).

References

[1] D. Bates, M. Maechler, B. Bolker. lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42, 2011. http://CRAN.R-project.org/package=lme4

[2] P. Gilbert. numDeriv: Accurate Numerical Derivatives. R package version 2012.3-1, 2012. http://CRAN.R-project.org/package=numDeriv

[3] L. A. Stefanski and D. D. Boos. The calculus of M-estimation. *The American Statistician*, 56(1):29–38, 2002.

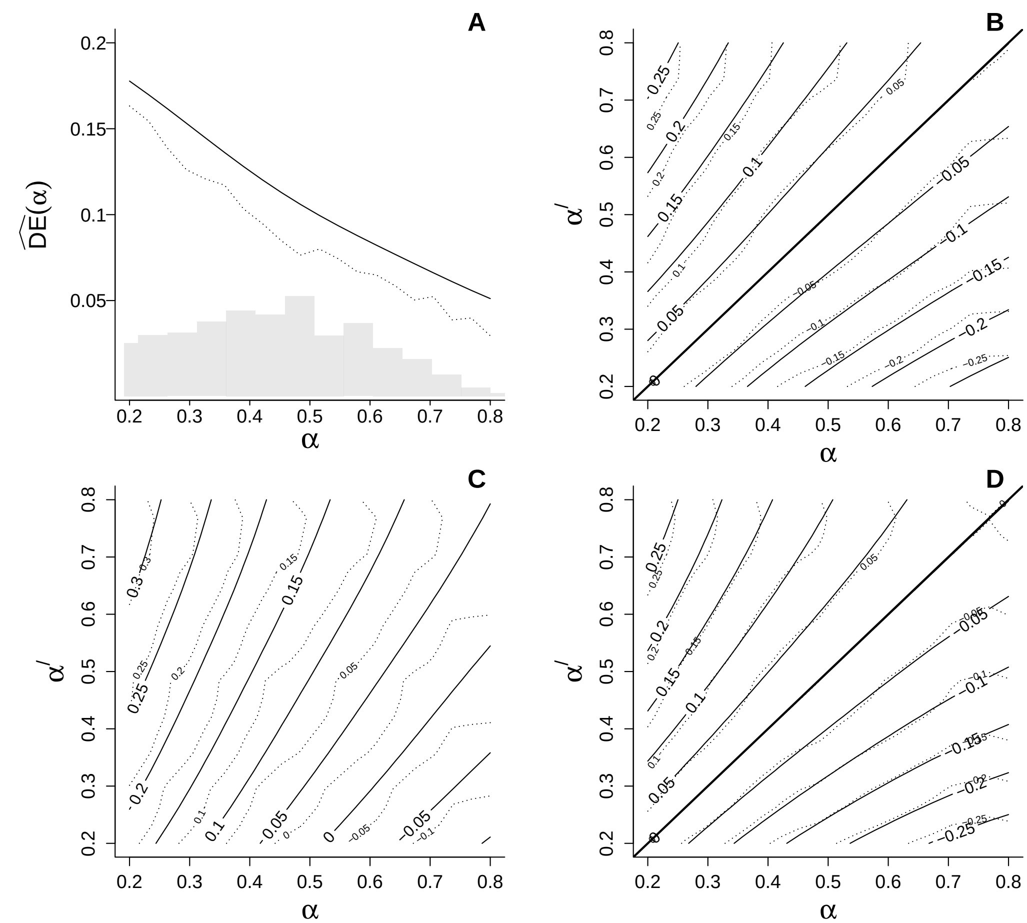
Web Tables

Web Table 1

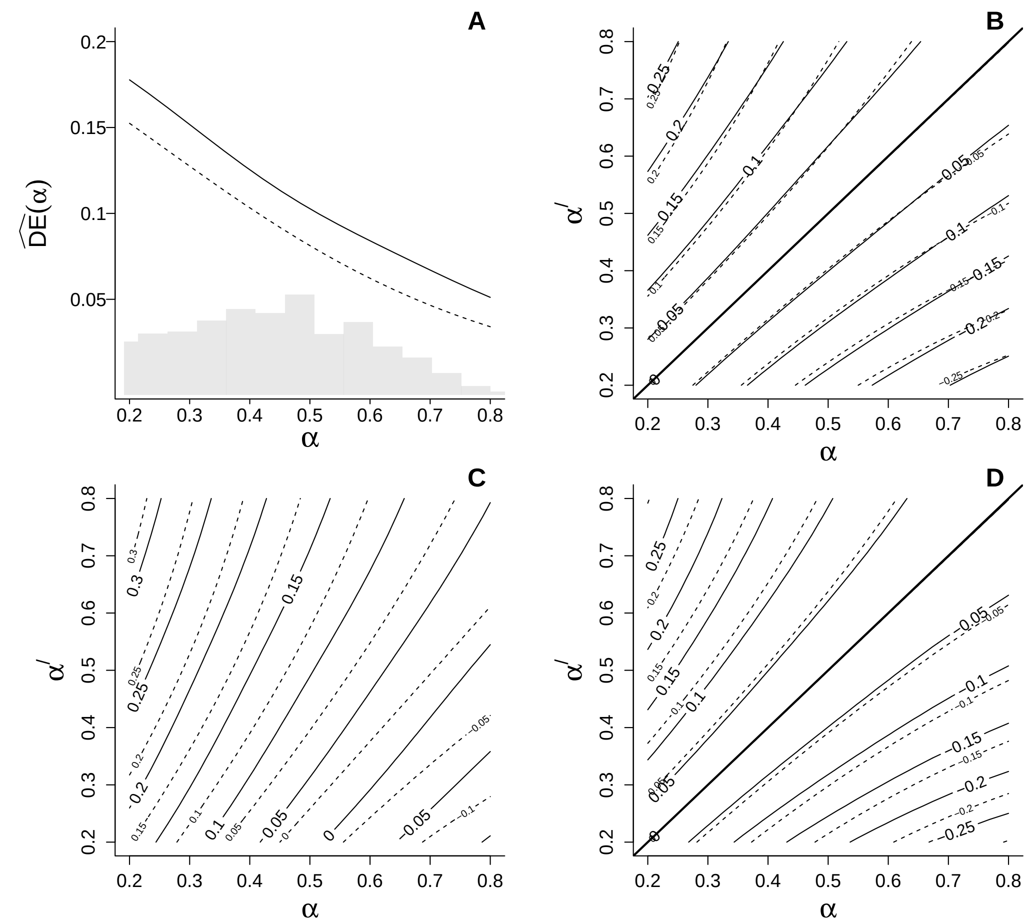
Results from simulation study described in Section 3 of the main paper. Truth *denotes the true value of the estimand;* IPW *denotes the average of the IPW estimates over the 1000 simulated data sets;* Bias = Truth−IPW *denotes the empirical bias;* ESE *denotes the empirical standard error;* ASE *is the average of the estimated sandwich-type standard errors; and* Cover *is the empirical coverage of Wald-type 95% confidence intervals*

Estimand	Truth	IPW	Bias	ESE	ASE	Cover
$\overline{DE}(0.3)$	0.152	0.152	0.000	0.013	0.014	97.4%
$\overline{IE}(0.3, 0.45)$	0.082	0.081	0.001	0.013	0.014	96.5%
$\overline{TE}(0.3, 0.45)$	0.196	0.195	0.001	0.015	0.017	96.7%
$\overline{OE}(0.3, 0.45)$	0.088	0.087	0.001	0.011	0.012	96.1%
$\overline{DE}(0.45)$	0.113	0.114	0.000	0.010	0.010	96.9%
$\overline{IE}(0.45, 0.6)$	0.067	0.064	0.003	0.011	0.011	95.5%
$\overline{TE}(0.45, 0.6)$	0.151	0.150	0.001	0.012	0.013	96.2%
$\overline{OE}(0.45, 0.6)$	0.066	0.065	0.002	0.008	0.008	96.0%
$\overline{DE}(0.6)$	0.084	0.086	-0.002	0.011	0.011	95.2%
$\overline{IE}(0.3, 0.6)$	0.149	0.145	0.004	0.018	0.019	95.9%
$\overline{TE}(0.3, 0.6)$	0.233	0.231	0.002	0.016	0.017	96.2%
$\overline{OE}(0.3, 0.6)$	0.154	0.151	0.003	0.014	0.015	95.5%

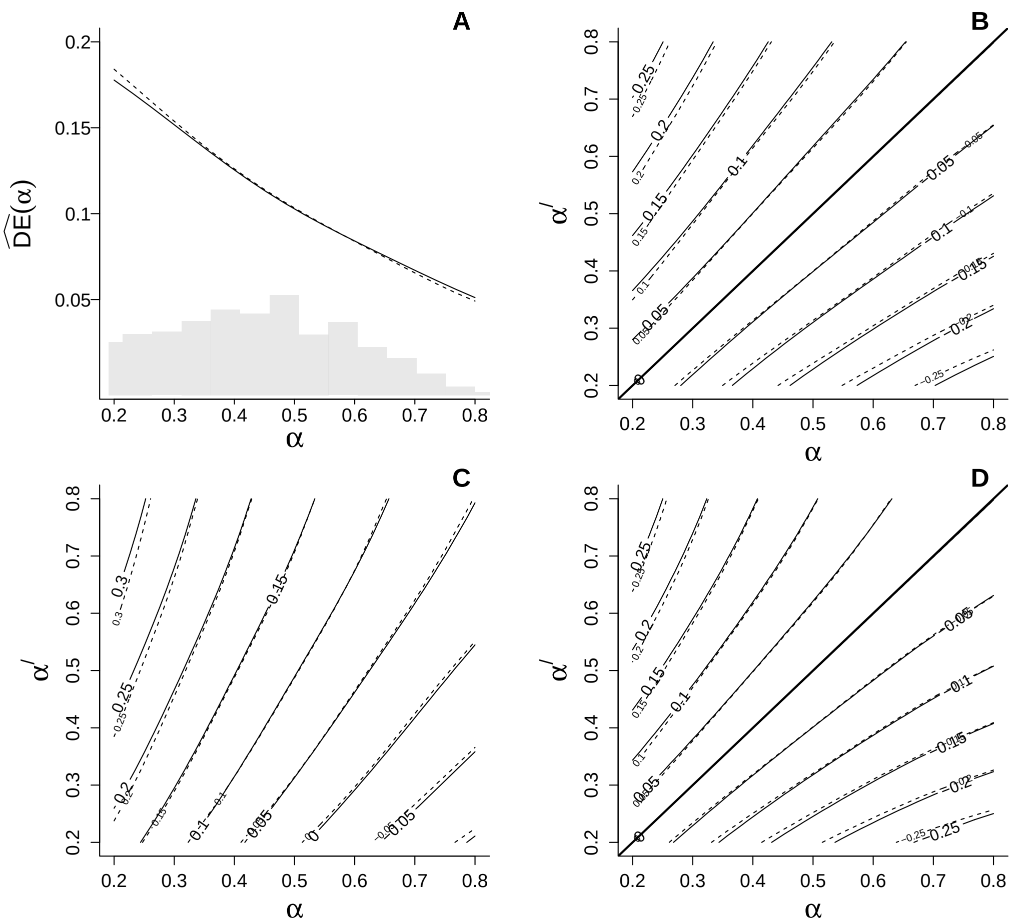
Web Figures



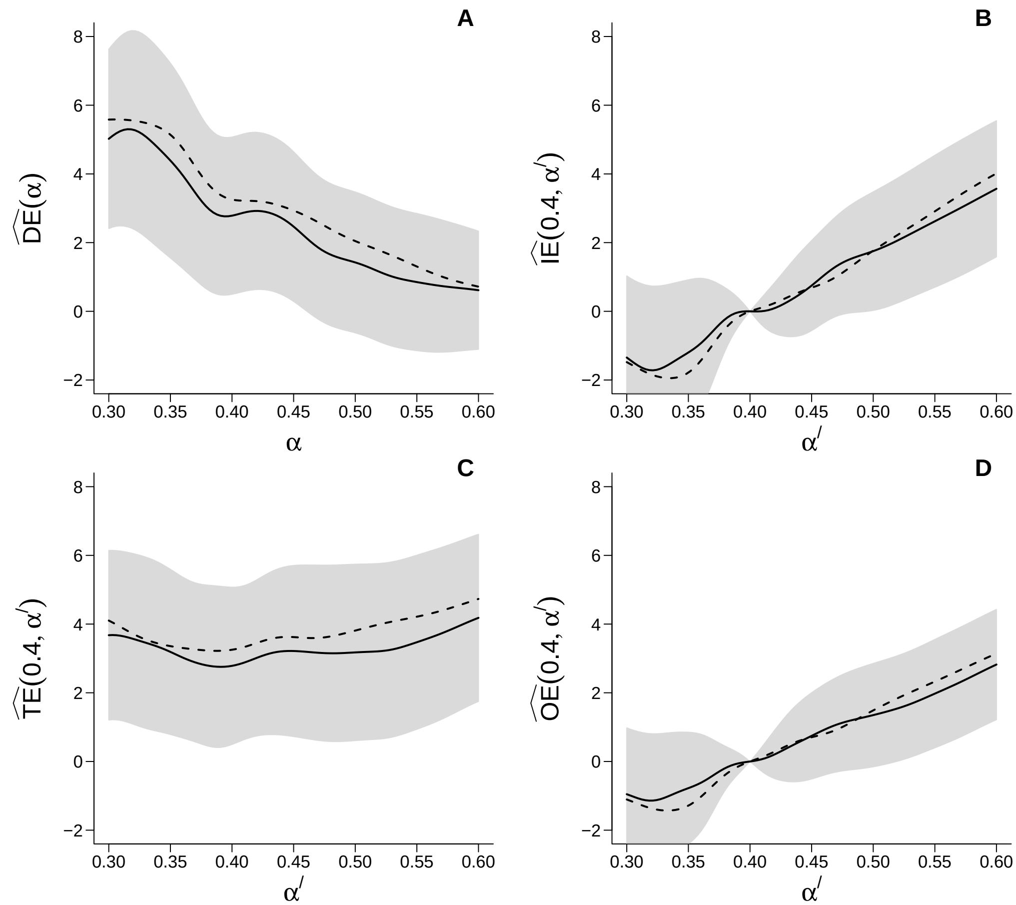
Web Figure 1: Naive estimates of (A) direct $\overline{DE}(\alpha)$, (B) indirect $\overline{IE}(\alpha, \alpha')$, (C) total $\overline{TE}(\alpha, \alpha')$, and (D) overall $\overline{OE}(\alpha, \alpha')$ effects from the simulation study. Solid lines represent the true effects, and dotted lines represent average effect estimates using the naive estimator described in Section 3 of the main paper. The histogram in (A) represents the distribution of vaccine coverage observed from the simulated data.



Web Figure 2: Outcome model-based estimates of (A) direct $\overline{DE}(\alpha)$, (B) indirect $\overline{IE}(\alpha, \alpha')$, (C) total $\overline{TE}(\alpha, \alpha')$, and (D) overall $\overline{OE}(\alpha, \alpha')$ effects from the simulation study. Solid lines represent the true effects, and dotted lines represent average effect estimates using the outcome model-based estimators described in Section 3. The histogram in (A) represents the distribution of vaccine coverage observed from the simulated data.



Web Figure 3: Misspecified IPW estimates of (A) direct $\overline{DE}(\alpha)$, (B) indirect $\overline{IE}(\alpha, \alpha')$, (C) total $\overline{TE}(\alpha, \alpha')$, and (D) overall $\overline{OE}(\alpha, \alpha')$ effects from the simulation study. Solid lines represent the true effects, and dashed lines represent average effect estimates using the IPW estimator with a misspecified propensity score model wherein the river distance covariate (X_{ij2}) was erroneously omitted. The histogram in (A) represents the distribution of vaccine coverage observed from the simulated data.



Web Figure 4: Sensitivity analysis of IPW estimates of (A) direct $\overline{DE}(\alpha)$, (B) indirect $\overline{IE}(0.40, \alpha')$, (C) total $\overline{TE}(0.40, \alpha')$, and (D) overall $\overline{OE}(0.40, \alpha')$ effects based on the cholera vaccine trial data. The solid line gives the estimates using the propensity score model that conditioned on age (linear and quadratic) and distance to the nearest river (linear and quadratic) as in Figures 3 and 4 of the main paper, and the gray regions around the effect estimates represent approximate pointwise 95% confidence intervals based on this propensity score model. The dashed lines correspond to IPW estimates based on an alternative propensity score model that conditioned on age (linear and quadratic), distance to the nearest river (linear and quadratic), distance to the nearest treatment center (linear), and religion.