Supporting Information for "Addressing Confounding and Exposure Measurement Error Using Conditional Score Functions" by Bryan S. Blette, Peter B. Gilbert, and Michael G. Hudgens

A   Large Sample Properties

In Web Appendix A, the large sample properties of the proposed estimators discussed in Section 3 of the main paper are proven.

A.1   G-formula CSM estimator

A.1.1   Large sample properties

Consistency and asymptotic normality of the g-formula estimator are proven using standard estimating equation theory (Stefanski and Boos, 2002). The original CSM estimator is an M-estimator and the g-formula can be written as a solution to an unbiased estimating equation. Thus, the proposed g-formula CSM estimator is an M-estimator with the estimating equation vector given in equation (3).

The parameter of interest $\mathrm{E}\{Y(\boldsymbol{a})\}$ appears in the last estimating equation of (3), and the proof of consistency and asymptotic normality of the g-formula CSM estimator relies on (i) the usual g-formula proof based on the causal assumptions made in Section 2 of

the main paper and (ii) that the CSM estimating equations were previously shown to be unbiased (Carroll et al., 2006). In particular, (ii) implies that $E[g^{-1}(\beta_0 + \boldsymbol{a}\beta_a + \boldsymbol{L}\beta_l + \boldsymbol{a}\beta_{al}\boldsymbol{L}^T)\}] = E\{E(Y|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{L})\}$ and then from (i) it follows that $E\{E(Y|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{L})\} = E[E\{Y(\boldsymbol{a})|\boldsymbol{L}\}] = E\{Y(\boldsymbol{a})\}$.

Therefore the estimating function for the parameter $E\{Y(\boldsymbol{a})\}$ is unbiased. Denote $\hat{\Theta}_{GF}$ as the solution to $\sum_{i=1}^{n} \psi_{GF-CSM}(Y_i, \boldsymbol{L}_i, \boldsymbol{A}_i^*, \Sigma_{me}, \hat{\Theta}_{GF}) = 0$. Then under certain regularity conditions $\sqrt{n}(\hat{\Theta}_{GF} - \Theta_{GF}) \rightarrow_d N(\boldsymbol{0}, A^{-1}B(A^{-1})^T)$ where A and B are consistently estimated by

$$\hat{A} = \frac{1}{n}\sum_{i=1}^{n} \frac{d}{d\Theta_{GF}^T}\psi_{GF-CSM}(Y_i, \boldsymbol{L}_i, \boldsymbol{A}_i^*, \Sigma_{me}, \hat{\Theta}_{GF})$$

$$\hat{B} = \frac{1}{n}\sum_{i=1}^{n} \psi_{GF-CSM}(Y_i, \boldsymbol{L}_i, \boldsymbol{A}_i^*, \Sigma_{me}, \hat{\Theta}_{GF})\psi_{GF-CSM}^T(Y_i, \boldsymbol{L}_i, \boldsymbol{A}_i^*, \Sigma_{me}, \hat{\Theta}_{GF})$$

A.1.2 Relationship to classical causal estimators

The CSM estimating equations reduce to the score equations for a generalized linear model (GLM) when the measurement error covariance matrix $\Sigma_{me} = 0_{m \times m}$ (Carroll et al., 2006). Thus under no measurement error, the procedure described above reduces to solving a stack of estimating equations corresponding to the common practice of using the g-formula while specifying a GLM for the outcome regression, making this standard procedure a special case of the proposed estimator.

A.2 IPW CSM estimator

A.2.1 Large sample properties

To show $E\left[\frac{f(\boldsymbol{A}^*)}{f_0(\boldsymbol{A}^*|\boldsymbol{L})}\{Y - E(Y|\boldsymbol{\Delta})\}\right] = 0$, we will show that $E\left(\frac{f(\boldsymbol{A}^*)}{f_0(\boldsymbol{A}^*|\boldsymbol{L})}Y\right) = E\left\{\frac{f(\boldsymbol{A}^*)}{f_0(\boldsymbol{A}^*|\boldsymbol{L})}E(Y|\boldsymbol{\Delta})\right\}$.

First,

$$E\left(\frac{f(\boldsymbol{A}^*)}{f_0(\boldsymbol{A}^*|\boldsymbol{L})}Y\right) = E\left\{E\left(\frac{f(\boldsymbol{A}^*)}{f_0(\boldsymbol{A}^*|\boldsymbol{L})}Y|L\right)\right\}$$

$$= \mathrm{E}\left\{\int_{a^*} \frac{f(a^*)}{f_0(a^*|L)}\mathrm{E}\left(Y|A^*=a^*,L\right)f_0(a^*|L)da^*\right\}$$

$$= \mathrm{E}\left[\int_{a^*} \mathrm{E}\left\{\mathrm{E}(Y|A^*=a^*,A,L)\right\}f(a^*)da^*\right]$$

$$= \mathrm{E}\left[\int_{a^*}\int_a E(Y|A^*=a^*,A=a,L)dF(a|L,A^*=a^*)f(a^*)da^*\right]$$

$$= \mathrm{E}\left[\int_{a^*}\int_a E\{Y(a)|L\}dF(a|L,A^*=a^*)f(a^*)da^*\right]$$

$$= \mathrm{E}\left[\mathrm{E}\{\mathrm{E}(Y|A=a,L)\}\cdot 1\right]$$

$$= \mathrm{E}\{Y(a)\} = \gamma_0 + a\gamma_a$$

where the fourth equality follows from the fact that Y is independent of A^* given A and L and the fifth equality follows from causal consistency and conditional exchangeability.

Now the second part reduces to

$$\mathrm{E}\left\{\frac{f(A^*)}{f_0(A^*|L)}\mathrm{E}(Y|\Delta)\right\} = \mathrm{E}\left[\int_{a^*}\mathrm{E}\left\{\mathrm{E}(Y|\Delta)|A^*=a^*,A=a,L\right\}f(a^*)da^*\right]$$

$$= \mathrm{E}\left[\mathrm{E}\{\mathrm{E}(Y|\Delta)|A=a,L\}\cdot 1\right]$$

$$= \mathrm{E}\left\{\frac{\gamma_0 + a\gamma_a + \mathrm{E}(\epsilon_{me}|A=a,L)\gamma_a + \gamma_a^T\Sigma_{me}\mathrm{E}(Y|A=a,L)\gamma_a/\sigma^2}{1 + \gamma_a^T\Sigma_{me}\gamma_a/\sigma^2}\right\}$$

$$= \frac{\gamma_0 + a\gamma_a + \gamma_a^T\Sigma_{me}\mathrm{E}[\mathrm{E}\{Y(a)|L\}]\gamma_a/\sigma^2}{1 + \gamma_a^T\Sigma_{me}\gamma_a/\sigma^2}$$

$$= \frac{\gamma_0 + a\gamma_a + \gamma_a^T\Sigma_{me}(\gamma_0 + a\gamma_a)\gamma_a/\sigma^2}{1 + \gamma_a^T\Sigma_{me}\gamma_a/\sigma^2}$$

$$= \frac{(\gamma_0 + a\gamma_a)(1 + \gamma_a^T\Sigma_{me}\gamma_a/\sigma^2)}{1 + \gamma_a^T\Sigma_{me}\gamma_a/\sigma^2}$$

$$= \gamma_0 + a\gamma_a$$

where the first equality repeats some steps from the other component. Thus, the proof is complete (for the intercept component).

Previous:

The estimating equation corresponding to the parameters of interest is

$$\sum_{i=1}^{n} \psi(Y_i, \boldsymbol{L}_i, \boldsymbol{A}_i^*, \Sigma_{me}, \Theta_{IPW}) = \sum_{i=1}^{n} SW_i(Y_i - \mathrm{E}[Y_i|\Delta_i])(1, \Delta_i)^T = 0$$

It suffices to show that the expectation of $\psi(Y, \boldsymbol{L}, \boldsymbol{A}^*, \Sigma_{me}, \Theta_{IPW})$ is equal to 0. For simplicity in writing integrals, suppose the exposure vector consists of only continuous-valued exposures. First consider a similar but infeasible estimator, with the same form but where it is weighted by the true propensity weights. Let $SW = dF(\boldsymbol{A})/dF_0(\boldsymbol{A}|\boldsymbol{L})$ where $F(\boldsymbol{A})$ is any distribution function and $F_0(\boldsymbol{A}|\boldsymbol{L})$ is the true conditional distribution function of the exposure $\boldsymbol{A}$ given confounders $\boldsymbol{L}$. Then

$$
\begin{aligned}
&\mathrm{E}\left[\frac{dF(\boldsymbol{A})}{dF_0(\boldsymbol{A}|\boldsymbol{L})}\{Y - \mathrm{E}(Y|\boldsymbol{\Delta})\}(1, \boldsymbol{\Delta})^T\right] \\
&= \mathrm{E}\left(\mathrm{E}\left[\frac{dF(\boldsymbol{A})}{dF_0(\boldsymbol{A}|\boldsymbol{L})}\{Y - \mathrm{E}(Y|\boldsymbol{\Delta})\}(1, \boldsymbol{\Delta})^T\Big|\boldsymbol{L}\right]\right) \\
&= \mathrm{E}\left[\int_{\boldsymbol{a}} \frac{dF(\boldsymbol{a})}{dF_0(\boldsymbol{a}|\boldsymbol{L})}\mathrm{E}[\{Y - \mathrm{E}(Y|\boldsymbol{\Delta})\}(1, \boldsymbol{\Delta})^T|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{L}]dF_0(\boldsymbol{a}|\boldsymbol{L})\right] \\
&= \mathrm{E}\left\{\int_{\boldsymbol{a}} \mathrm{E}[\{Y - \mathrm{E}(Y|\boldsymbol{\Delta})\}(1, \boldsymbol{\Delta})^T|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{L}]dF(\boldsymbol{a})\right\} \\
&= \int_{\boldsymbol{a}} \mathrm{E}(\mathrm{E}[\{Y - \mathrm{E}(Y|\boldsymbol{\Delta})\}|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{L}, \boldsymbol{\Delta}](1, \boldsymbol{\Delta})^T)dF(\boldsymbol{a}) \\
&= \int_{\boldsymbol{a}} \mathrm{E}[\{\mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta}) - \mathrm{E}(Y|\boldsymbol{\Delta})\}(1, \boldsymbol{\Delta})^T]dF(\boldsymbol{a}) \\
&= \int_{\boldsymbol{a}} \{\mathrm{E}(Y|\boldsymbol{\Delta}) - \mathrm{E}(Y|\boldsymbol{\Delta})\}(1, \boldsymbol{\Delta})^T dF(\boldsymbol{a}) \\
&= 0
\end{aligned}
$$

where the 5th equality follows because Y is independent of $\boldsymbol{A}$ conditional on $\boldsymbol{\Delta}$ and $\boldsymbol{L}$. Thus the infeasible estimator is consistent and asymptotically normal by standard M-estimator theory.

This estimator is infeasible because $\boldsymbol{A}$ is not observed and $F_0(\boldsymbol{A}|\boldsymbol{L})$ is not known, so instead a plug-in type estimator is used for SW. Let $F(\boldsymbol{a}^*; \xi)$ and $F(\boldsymbol{a}^*|\boldsymbol{L}; \zeta)$ be finite-dimensional parametric models of the distributions of $\boldsymbol{A}$ and of $\boldsymbol{A}|\boldsymbol{L}$, let $(\hat{\xi}, \hat{\zeta})$ denote some consistent estimator of (ξ, ζ), and assume $(\hat{\xi}, \hat{\zeta})$ converges in probability to some constant vector (ξ_0, ζ_0). Note these parametric models need not be correctly specified, but instead we make the following assumption:

$$E\{dF(\boldsymbol{A}^*; \xi_0)/dF(\boldsymbol{A}^*|\boldsymbol{L}; \zeta_0) \mid Y, \boldsymbol{\Delta}, \boldsymbol{L}\} = E\{dF(\boldsymbol{A})/dF_0(\boldsymbol{A}|\boldsymbol{L}) \mid Y, \boldsymbol{\Delta}, \boldsymbol{L}\}.$$

Under this assumption, it follows from the infeasible estimator proof above that the IPW CSM estimator computing using $SW = dF(\boldsymbol{A}^*; \hat{\xi})/dF(\boldsymbol{A}^*; \hat{\zeta})$ is also the solution to a set unbiased estimating equations and therefore is consistent and asymptotically normal. The estimator of the asymptotic variance of the IPW CSM estimator is given by the usual sandwich estimator as described in the previous section.

A.2.2 Relationship to classical causal estimators

If the measurement error covariance matrix $\Sigma_{me} = \boldsymbol{0}_{m \times m}$, then $\boldsymbol{A}^* = \boldsymbol{A}$ and the sufficient statistic $\boldsymbol{\Delta}$ reduces to the observed exposure vector. Then the IPW estimator reduces to the form $\sum_{i=1}^{n} \psi(Y_i, \boldsymbol{L}_i, \boldsymbol{A}_i^*, \Theta_{IPW}) = 0$ where:

$$\psi(Y, \boldsymbol{L}, \boldsymbol{A}^*, \Theta_{IPW}) = \begin{bmatrix} \psi_{PS}(\boldsymbol{L}, \boldsymbol{A}^*, \zeta) \\ SW\{Y - \mathrm{E}(Y|\boldsymbol{A}^*)\}(1, \boldsymbol{A}^*)^T \\ SW\left[\phi - \frac{\{Y - \mathrm{E}(Y|\boldsymbol{A}^*)\}^2}{Var(Y|\boldsymbol{A}^*)/\phi}\right] \end{bmatrix}$$

This is exactly the score function vector for a GLM weighted by SW. Thus, an IPW estimator fit using a weighted GLM for outcome Y is a special case of the proposed IPW CSM estimator where there is no measurement error present.

A.3 DR CSM Estimator

A.3.1 Large Sample Properties

Once again, consistency and asymptotic normality of the proposed estimator is proven using M-estimator theory. Recall from Section 3.4 that this estimator is a solution to the estimating equation $\sum_{i=1}^{n} \psi_{DR-CSM}(Y_i, \boldsymbol{L}_i, \boldsymbol{A}_i^*, \Sigma_{me}, \Theta_{DR}) = 0$, where $\Theta_{DR} = (\zeta, \Theta_{GF})$ and

$$\psi_{DR-CSM}(Y, \boldsymbol{L}, \boldsymbol{A}^*, \Sigma_{me}, \Theta_{DR}) = \begin{bmatrix} \psi_{PS}(\boldsymbol{L}, \boldsymbol{A}^*, \zeta) \\ SW\{Y - \mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta})\}(1, \boldsymbol{L}, \boldsymbol{\Delta}, \boldsymbol{L} \otimes \boldsymbol{\Delta})^T \\ SW[\phi - \{Y - \mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta})\}^2/\{\mathrm{Var}(Y|\boldsymbol{L}, \boldsymbol{\Delta})/\phi\}] \\ \beta_0 + \boldsymbol{a}\beta_a + \boldsymbol{L}\beta_l + \boldsymbol{a}\beta_{al}\boldsymbol{L}^T - \mathrm{E}\{Y(\boldsymbol{a})\} \end{bmatrix}$$

First, suppose that the outcome regression is correctly specified. Then $\mathrm{E}(\beta_0 + \boldsymbol{a}\beta_a + \boldsymbol{L}\beta_l + \boldsymbol{a}\beta_{al}\boldsymbol{L}^T) = \mathrm{E}\{\mathrm{E}(Y|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{L})\} = \mathrm{E}[\mathrm{E}\{Y(\boldsymbol{a})|\boldsymbol{L}\}] = \mathrm{E}\{Y(\boldsymbol{a})\}$ such that $\mathrm{E}[(\beta_0 + \boldsymbol{a}\beta_a + \boldsymbol{L}\beta_l + \boldsymbol{a}\beta_{al}\boldsymbol{L}^T) - \mathrm{E}\{Y(\boldsymbol{a})\}] = 0$.

Now suppose instead that the propensity score model(s) are correctly specified and that $dF(\boldsymbol{A}|\boldsymbol{L})/d\hat{F}(\boldsymbol{A}^*|\boldsymbol{L})$ does not depend on $\boldsymbol{L}$ as described in the IPW proof. For simplicity in writing integrals, suppose all exposures are continuous and that the stabilized weights are given by a ratio of densities. Recall that under the identity link,

$$\boldsymbol{\Delta} = \boldsymbol{A}^* + (\beta_a + \boldsymbol{L}\beta_{al})\Sigma_{me}\boldsymbol{Y}/\sigma^2$$

$$\mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta}) = \frac{\beta_0 + \boldsymbol{L}\beta_l + \boldsymbol{\Delta} \cdot (\beta_a + \boldsymbol{L}\beta_{al})}{1 + (\beta_a + \boldsymbol{L}\beta_{al})\Sigma_{me}(\beta_a + \boldsymbol{L}\beta_{al})^T/\sigma^2}$$

Temporary note to self: Goal is to do the same as the non-ME proof, i.e., use the fact that the second EE of ψ_{DR-CSM} is equal to 0 to show that the fourth EE is equal to 0. So need to simplify the second EE such that we get an $E\{Y(a)\}$ and a $E(\beta_0 + \boldsymbol{a}\beta_a + \boldsymbol{L}\beta_l + \boldsymbol{a}\beta_{al}\boldsymbol{L}^T)$ towards the end. Before we utilized $E(Y|\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{\Delta}) = E(Y|\boldsymbol{L}, \boldsymbol{\Delta})$. Now we utilize $E(Y|\boldsymbol{A}, \boldsymbol{L}, \boldsymbol{\Delta}) = E(Y|\boldsymbol{A}, \boldsymbol{L})$ instead.

Consider two elements of the second estimating equation of ψ_{DR-CSM}, and define β^* such that

$$
\begin{aligned}
0 &= \mathrm{E}\left[\frac{d\hat{F}(\boldsymbol{A}^*)}{d\hat{F}(\boldsymbol{A}^*|\boldsymbol{L})}\{Y - \mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta}; \beta^*)\}(1, \boldsymbol{\Delta})^T\right] \\
&= \mathrm{E}\left(\mathrm{E}\left[\frac{d\hat{F}(\boldsymbol{A}^*)}{d\hat{F}(\boldsymbol{A}^*|\boldsymbol{L})}\{Y - \mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta}; \beta^*)\}(1, \boldsymbol{\Delta})^T\bigg|\boldsymbol{L}\right]\right) \\
&= \mathrm{E}\left[\int_{\boldsymbol{a}} \frac{d\hat{F}(\boldsymbol{a} + \epsilon_{me})}{d\hat{F}(\boldsymbol{a} + \epsilon_{me}|\boldsymbol{L})}\mathrm{E}[\{Y - \mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta}; \beta^*)\}(1, \boldsymbol{\Delta})^T|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{L}]dF_0(\boldsymbol{a}|\boldsymbol{L})\right] \\
&= \mathrm{E}\left\{\int_{\boldsymbol{a}} \frac{dF_0(\boldsymbol{a}|\boldsymbol{L})}{d\hat{F}(\boldsymbol{a} + \epsilon_{me}|\boldsymbol{L})}\mathrm{E}[\{Y - \mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta}; \beta^*)\}(1, \boldsymbol{\Delta})^T|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{L}]d\hat{F}(\boldsymbol{a} + \epsilon_{me})\right\} \\
&= \int_{\boldsymbol{a}} \frac{dF_0(\boldsymbol{a}|\boldsymbol{L})}{d\hat{F}(\boldsymbol{a} + \epsilon_{me}|\boldsymbol{L})}\mathrm{E}(\mathrm{E}[\{Y - \mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta}; \beta^*)\}|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{L}, \boldsymbol{\Delta}](1, \boldsymbol{\Delta})^T)d\hat{F}(\boldsymbol{a} + \epsilon_{me}) \\
&= \int_{\boldsymbol{a}} \frac{dF_0(\boldsymbol{a}|\boldsymbol{L})}{d\hat{F}(\boldsymbol{a} + \epsilon_{me}|\boldsymbol{L})}\mathrm{E}\{\mathrm{E}(Y|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{L}) - \mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta}; \beta^*)\}(1, \boldsymbol{\Delta})^T d\hat{F}(\boldsymbol{a} + \epsilon_{me}) \\
&= \int_{\boldsymbol{a}} \frac{dF_0(\boldsymbol{a}|\boldsymbol{L})}{d\hat{F}(\boldsymbol{a} + \epsilon_{me}|\boldsymbol{L})}[\mathrm{E}\{Y(a)\} - \mathrm{E}\{\mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta}; \beta^*)|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{L}\}](1, \boldsymbol{\Delta})^T d\hat{F}(\boldsymbol{a} + \epsilon_{me})
\end{aligned}
$$

Now simplify

$$
\begin{aligned}
&\mathrm{E}\{\mathrm{E}(Y|\mathbf{L}, \Delta; \beta^*)|\mathbf{A} = \mathbf{a}, \mathbf{L}\} \\
&= \frac{\beta_0^* + \mathbf{L}\beta_l^* + \{\mathbf{a} + E(\epsilon_{me}|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{L}) + (\beta_a^* + \mathbf{L}\beta_{al}^*)\Sigma_{me}E(\mathbf{Y}|\boldsymbol{A} = \boldsymbol{a}, \boldsymbol{L})/\sigma^2\} \cdot (\beta_a^* + \mathbf{L}\beta_{al}^*)}{1 + (\beta_a^* + \mathbf{L}\beta_{al}^*)\Sigma_{me}(\beta_a^* + \mathbf{L}\beta_{al}^*)^T/\sigma^2} \\
&= \frac{\beta_0^* + \mathbf{L}\beta_l^* + \mathbf{a}\beta_a^* + \mathbf{a}\beta_{al}^*\mathbf{L}^T + ((\beta_a^* + \mathbf{L}\beta_{al}^*)\Sigma_{me}(\beta_0^* + \mathbf{L}\beta_l^* + \mathbf{a}\beta_a^* + \mathbf{a}\beta_{al}^*\mathbf{L}^T)/\sigma^2) \cdot (\beta_a^* + \mathbf{L}\beta_{al}^*)}{1 + (\beta_a^* + \mathbf{L}\beta_{al}^*)\Sigma_{me}(\beta_a^* + \mathbf{L}\beta_{al}^*)^T/\sigma^2} \\
&= \frac{(\beta_0^* + \mathbf{L}\beta_l^* + \mathbf{a}\beta_a^* + \mathbf{a}\beta_{al}^*\mathbf{L}^T) \cdot \{1 + (\beta_a^* + \mathbf{L}\beta_{al}^*)\Sigma_{me}(\beta_a^* + \mathbf{L}\beta_{al}^*)^T/\sigma^2\}}{1 + (\beta_a^* + \mathbf{L}\beta_{al}^*)\Sigma_{me}(\beta_a^* + \mathbf{L}\beta_{al}^*)^T/\sigma^2} \\
&= \beta_0^* + \mathbf{L}\beta_l^* + \mathbf{a}\beta_a^* + \mathbf{a}\beta_{al}^*\mathbf{L}^T
\end{aligned}
$$

So

$$
0 = \int_{\boldsymbol{a}} \frac{dF_0(\boldsymbol{a}|\boldsymbol{L})}{d\hat{F}(\boldsymbol{a}^* + \epsilon_{me}|\boldsymbol{L})}[\mathrm{E}\{Y(a)\} - (\beta_0^* + \mathbf{L}\beta_l^* + \mathbf{a}\beta_a^* + \mathbf{a}\beta_{al}^*\mathbf{L}^T)](1, \boldsymbol{\Delta})^T d\hat{F}(\boldsymbol{a} + \epsilon_{me})
$$

7

This doesn't seem quite the same as the non-ME setting. Did I lose track of an expectation somewhere?

...Thus, the DR-CSM estimator is consistent and asymptotically normal as long as one of the propensity and outcome models is correctly specified. The estimator of the asymptotic variance is given by the usual sandwich variance estimator.

A.3.2 Relationship to classical causal estimators

Once again, under no measurement error the CSM equations reduce to the score equations of a GLM (here a linear model). Thus the proposed DR CSM estimator will reduce to the DR estimator described in Kang and Schafer (2007) where the specified outcome regression is a linear model.

A.3.3 Alternative DR estimator

A second DR estimator, similar to those described in Robins (2000) and Neugebauer and van der Laan (2005), is given by the solution to the set of estimating equations $\sum_{i=1}^{n} \psi_{DR-CSM-alt}(Y_i, \boldsymbol{L}_i, \boldsymbol{A}_i^*, \Theta_{DR-alt}) = 0$ where $\Theta_{DR-alt} = \Theta_{IPW}$ and

$$\psi_{DR-CSM-alt}(Y, \boldsymbol{L}, \boldsymbol{A}^*, \Theta_{DR-alt}) = \begin{bmatrix} SW\{Y - \mathrm{E}(Y|\boldsymbol{\Delta}) - Q(\boldsymbol{A}^*, \boldsymbol{L})\} + \int_{\boldsymbol{a}} dF(\boldsymbol{a})Q(\boldsymbol{a}, \boldsymbol{L})d\mu(\boldsymbol{a}) \\ SW\{Y - \mathrm{E}(Y|\boldsymbol{\Delta}) - Q(\boldsymbol{A}^*, \boldsymbol{L})\}\boldsymbol{\Delta} + \int_{\boldsymbol{a}} dF(\boldsymbol{a})Q(\boldsymbol{a}, \boldsymbol{L})d\mu(\boldsymbol{a}) \end{bmatrix},$$

$dF(\boldsymbol{a})$ is the numerator of the stabilized weights, $d\mu(\boldsymbol{a})$ is an appropriate measure (Lebesgue if all exposures are continuous), and $Q(\boldsymbol{A}, \boldsymbol{L}) = \beta_0 + \boldsymbol{A}\beta_a + \boldsymbol{L}\beta_l + \boldsymbol{A}\beta_{al}\boldsymbol{L}^T - \gamma(1, \boldsymbol{A})^T$ when the outcome regression is specified in a similar way to that used for the DR CSM estimator (main effects for exposures and confounders, as well as relevant interactions). Note the second DR estimator targets the same MSM parameter estimand as the IPW estimator and is consistent if either (i) the MSM and propensity model are correctly specified or (ii) the MSM and outcome model are correctly specified. The MSM and outcome model should be compatible; these compatible scenarios should largely line up with the scenarios where the DR CSM estimator is used, such as when both the MSM and outcome model use an identity

link. However, the DR CSM estimator is much easier to calculate in practice.

A.4   Uniqueness of EE solutions

The proofs above showing consistency and asymptotic normality of the proposed estimators require each set of estimating equations to have a unique solution. Prior work has shown that similar conditional score equations do not always have unique solutions, but that multiple solutions are very rare in practice (Stefanski and Carroll, 1987). In the various simulations of this paper, multiple solutions or estimator divergence were encountered with similar rarity, at most 1 or 2 times per 1000 simulations, unless considering extreme data generating mechanisms. Thus, the estimators should have good behavior in general, but practitioners should be aware of rare instances of multiple solutions, unusual estimates, and/or root-solving algorithm divergence errors.

B   Two-Phase Sampling

B.1   Two-phase sampling method

Many studies (including the HVTN 505 trial) use a two-phase sampling design. Such a design is particularly useful when the primary exposure(s) and outcome are easy to measure, but exposures and covariates of secondary interest are expensive or difficult to measure. Because each of the proposed methods above belongs to the estimating equation framework, it is straightforward to incorporate previously described methods for causal inference from studies with two-phase sampling. In this section, one such approach is demonstrated using a simulation study. In particular, for this simulation and the application section analysis, a simple inverse probability of sampling weights method is used (Wang et al., 2009).

The method is implemented by weighting each individual's contribution to the estimating equations by the inverse probability of selection for the second-phase of the study (multiplying treatment weights by sampling weights for the IPW-CSM and DR-CSM estimators)

and restricting the analysis to those selected. This method is a good choice for the subset of the HVTN 505 trial that is the focus of Section 5 of the main paper, particularly because all exposures of interest were measured in the second-phase sub-sample and no exposures were measured in the full sample.

B.2 Two-phase sampling simulations

The structure of the first simulation study described in Section 4 of the main paper is replicated, but under a two-phase sampling design. In particular, a case-cohort design is used where the exposure is measured for a random sub-cohort as well as for every case. This is done for a sample size of $n = 2000$ under three scenarios, with sub-cohorts of size 5%, 10% and 25%. The results of 2000 simulation runs are presented in Web Figure 1 and Web Table 1.

The methods perform similar to the full-sampling simulation provided in the main paper, although there is some bias and under-coverage when the sub-cohorts are smaller, likely due to a low effective sample size. In addition, the estimators failed to converge in some of the small sub-cohort settings. However, the DR CSM estimator with sampling weights converged in all analyses presented in Section 5 of the main paper.

C   Additional Simulations

In this section, additional simulations are presented under two assumption violations: (i) when positivity doesn't hold and (ii) when measurement error doesn't follow a classical additive model.

C.1 Positivity violation

To evaluate the proposed g-formula CSM method under positivity violation, the general structure of the first simulation study from Section 4 of the main paper is replicated almost

10

exactly. A moderate positivity violation is created by changing how the treatment A_1 is generated from $\mathcal{N}(2 + 0.3L_1 - 0.5L_2, 0.6)$ to $\mathcal{N}(2 + 0.3L_1 - 0.5L_2, 0.35)$. This breaks the phenomenon of mostly overlapping treatment values experienced by simulated subjects with different covariate values, although in a technical sense is not a structural violation of positivity since the distributions would have the same support given infinite sample size. The results of the simulation study are presented in Web Figure 2 and Web Table 2, following the same format as Table 1 and Figure 2 in the main paper.

The results overall look similar to that in Table 1 of the main paper. There is some bias and undercoverage for the proposed g-formula CSM methods, but the proposed method still generally performs better than the comparator methods in this scenario. Some bias at the extremes of the exposure support is expected and not fully due to the positivity violation, given that there is very little data at the extremes with the new data generating mechanism. A reasonable range to evaluate the methods would be from 0.5 to 3.5 in these simulations.

Positivity violations become more likely with more treatment variables and with treatment variables that are continuous or take on many values. In these settings positivity should receive just as much scrutiny as the conditional exchangeability assumption. If positivity is implausible, it may be possible to define an estimator in our setting similar to that described in Neugebauer and van der Laan (2005) which was robust to their analogous "experimental treatment assumption".

C.2 Non-additive measurement error

Next the proposed methods are evaluated when treatment measurement error does not follow the classical additive model. In particular, the second simulation study from Section 4 of the main paper is replicated, but the simulation of mismeasured treatment A_3^* is changed such that it follows a multiplicative error model simulated as $A_3^* = A_3 \epsilon_{me_1}$ where $\epsilon_{me_1} \sim \mathcal{N}(1, 0.1)$. The methods are still performed assuming additive measurement error with known measurement error covariance as specified in section 4 of the main paper, and the A_3^* distribution

11

under this additive ME assumption is similar to the distribution under the true multiplicative ME generative model. The results are presented in Web Table 3. The proposed IPW CSM method continued to perform well for treatments A_1 and A_2 for which the assumptions hold, but exhibited strong bias for the A_3 effect. Practitioners of the proposed methods should be cautious that if classical additive measurement error models do not hold for their exposures, they may get worse results than even standard regression models.

D   More complex model specifications for the IPW CSM estimator

The proposed IPW CSM estimator assumes a linear marginal structural model form. While this is helpful to match the conditional score framework described in Section 3.1, it may be too restrictive for some potential applications. To this end we note that transformations of elements of $\boldsymbol{A}$ and interactions thereof can be included in the MSM specification as long as they are either assumed to be correctly measured or assumed to follow a classical additive measurement error model. For example, if an exposure is strictly positive, then its log transform may be assumed to follow an additive measurement error model and can be included in the model. In general, transformations of correctly measured exposures can be included in the MSM specification without restriction.

Finally, while conditional score functions are somewhat limited in scope in terms of model specification, the related method of corrected score functions has been extended to problems of additive but non-normal measurement error (Buzas and Stefanski, 1996) and to non-additive measurement models in certain cases (Nakamura, 1990; Li, Palta, and Shao, 2004). Describing how to use such corrected score functions to estimate causal parameters could be the focus of future work in this area.

Web Table 1. *Simulation study for case-cohort design. Bias: 100 times the average bias across simulated data sets for each method; ASE: 100 times the average of estimated standard errors; ESE: 100 times the standard deviation of parameter estimates; Cov: Empirical coverage of 95% confidence intervals for each method, rounded to the nearest integer. % FTC: Percent of simulations which failed to converge, rounded to the nearest integer.*

Sub-cohort Size	Bias	ASE	ESE	Cov	% FTC
5%	4.1	8.7	7.1	84%	6%
10%	2.4	6.3	5.6	90%	2%
25%	0.9	4.1	3.9	94%	0%

Web Table 2. *Simulation study under positivity violation. Bias, ASE, ESE, and Cov defined as in Web Table 1.*

Estimator	Bias	ASE	ESE	Cov
Regression	-69.8	21.5	21.8	11%
CSM	-17.0	80.3	68.5	95%
G-formula	-6.3	3.0	3.0	44%
G-formula CSM	2.2	9.3	9.1	92%

Web Table 3. *Simulation study under non-additive measurement error. Bias, ASE, ESE, and Cov defined as in Web Table 1.*

Parameter	Estimator	Bias	ASE	ESE	Cov
γ_1	Regression	4.9	14.0	13.3	93%
γ_1	CSM	21.7	20.0	19.3	82%
γ_1	IPW	-9.9	9.1	9.0	79%
γ_1	IPW CSM	0.6	13.0	12.7	95%
γ_2	Regression	10.3	28.0	27.7	93%
γ_2	CSM	8.7	29.4	28.4	93%
γ_2	IPW	0.0	20.0	19.7	94%
γ_2	IPW CSM	-0.9	20.4	20.1	95%
γ_3	Regression	1.9	15.8	15.4	94%
γ_3	CSM	-35.0	33.5	32.7	86%
γ_3	IPW	4.7	15.9	15.5	93%
γ_3	IPW CSM	-29.0	33.0	32.0	88%
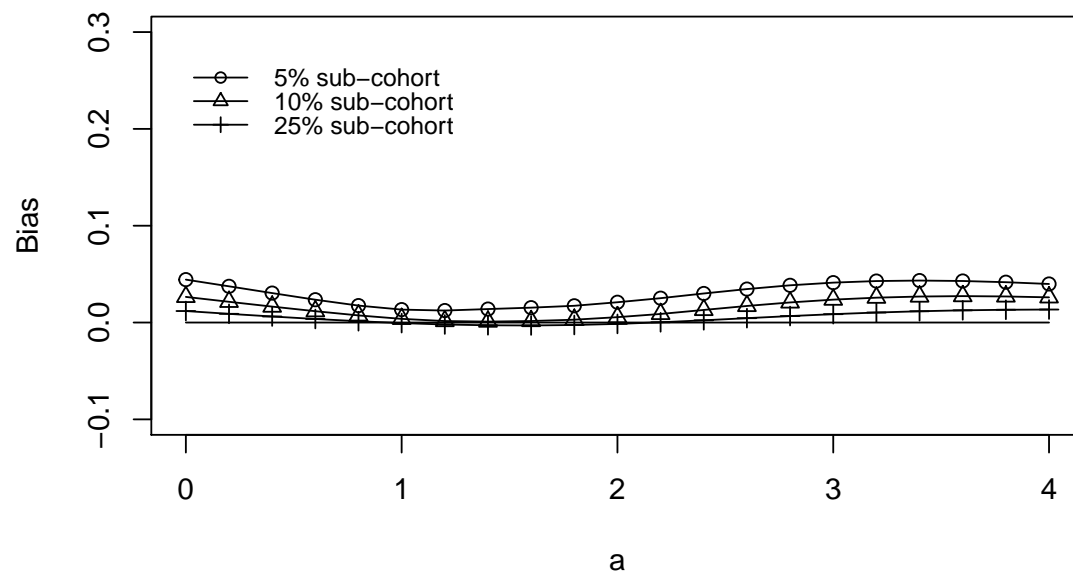
Web Figure 1: Estimated dose-response curve bias for the DR CSM method compared across three sub-cohort sizes. Bias refers to the average bias across 2,000 simulated data sets for each method evaluated at each point on the horizontal axis $a = (0, 0.2, 0.4, ..., 4)$.
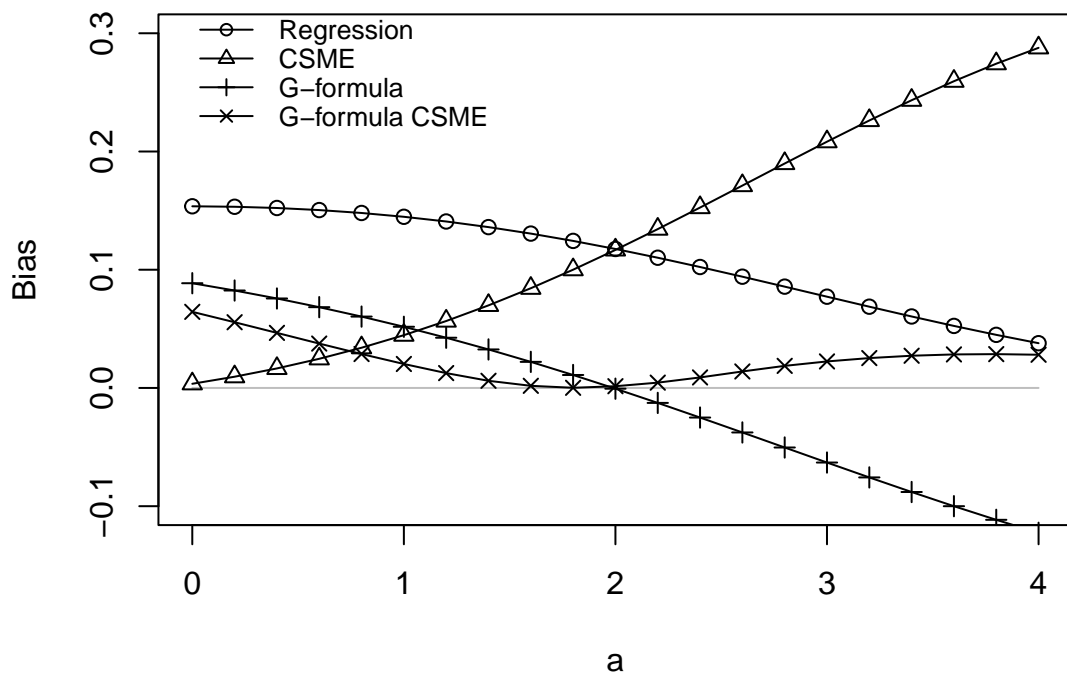
Web Figure 2: Estimated dose-response curve bias for each of the four methods under positivity violation. Bias refers to the average bias across 2,000 simulated data sets for each method evaluated at each point on the horizontal axis $a = (0, 0.2, 0.4, ..., 4)$.

References

Buzas, J. and Stefanski, L. (1996). A note on corrected-score estimation. *Statistics & Probability Letters* **28,** 1–8.

Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective.* CRC press.

Kang, J. D. and Schafer, J. L. (2007). Demystifying double robustness: A comparison of alternative strategies for estimating a population mean from incomplete data. *Statistical Science* **22,** 523–539.

Li, L., Palta, M., and Shao, J. (2004). A measurement error model with a Poisson distributed surrogate. *Statistics in Medicine* **23,** 2527–2536.

Nakamura, T. (1990). Corrected score function for errors-in-variables models: Methodology and application to generalized linear models. *Biometrika* **77,** 127–137.

Neugebauer, R. and van der Laan, M. (2005). Why prefer double robust estimators in causal inference? *Journal of Statistical Planning and Inference* **129,** 405–426.

Robins, J. M. (2000). Marginal structural models versus structural nested models as tools for causal inference. In Halloran, M. and Berry, D., editors, *Statistical Models in Epidemiology, the Environment, and Clinical Trials. The IMA Volumes in Mathematics and its Applications, vol 116.*, pages 95–133. Springer, New York, NY.

Rosenbaum, P. R. (1987). Model-based direct adjustment. *Journal of the American Statistical Association* **82,** 387–394.

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

Stefanski, L. A. and Carroll, R. J. (1987). Conditional scores and optimal scores for generalized linear measurement-error models. *Biometrika* **74,** 703–716.

Wang, W., Scharfstein, D., Tan, Z., and MacKenzie, E. J. (2009). Causal inference in outcome-dependent two-phase sampling designs. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71,** 947–969.