

# Web Appendix for “Addressing concurrent confounding and exposure measurement error using a conditional score approach”

Bryan S. Blette, Peter B. Gilbert, Michael G. Hudgens

## **1 Web Appendix A: Large Sample Properties**

In Web Appendix A, the large sample properties of the proposed estimators discussed in section 3.6 of the paper are proven.

### **1.1 G-formula CSME estimator**

#### **1.1.1 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 CSME estimator is an M-estimator and the g-formula can be written in the form of an unbiased estimating equation (EE). Thus, the proposed g-formula CSME estimator can be written as

$\sum_{i=1}^n \psi_{GF-CSME}(Y_i, \mathbf{L}_i, \mathbf{A}_i^*, \Sigma, \Theta) = 0$ , where  $\Theta = (\beta_0, \beta_a^T, \beta_l^T, \phi, \psi_1, \dots, \psi_m)$  and:

$$\psi_{GF-CSME}(Y_i, \mathbf{L}_i, \mathbf{A}_i^*, \Sigma_{me}, \Theta) = \begin{bmatrix} \{Y_i - E(Y_i | \mathbf{L}_i, \Delta_i)\} (1, \mathbf{L}_i, \Delta_i, \mathbf{L}_i \otimes \Delta_i)^T \\ \phi - \{Y_i - E(Y_i | \mathbf{L}_i, \Delta_i)\}^2 / \{\text{Var}(Y_i | \mathbf{L}_i, \Delta_i) / \phi\} \\ \beta_{a1} \mathbf{l}_i^T + \beta_{a1} - \gamma_1 \\ \vdots \\ \beta_{am} \mathbf{l}_i^T + \beta_{am} - \gamma_m \end{bmatrix}$$

Each of the final  $m$  EEs in the stack is an algebraic simplification of the g-formula in this context, and the EE for an arbitrary parameter  $\psi_k$  can be shown to have expectation 0 as follows. First note that under the three identification assumptions of causal consistency, conditional exchangeability, and positivity,  $E\{Y(\mathbf{a})\} = \frac{1}{n} \sum_i E(Y | \mathbf{A} = \mathbf{a}, \mathbf{L}_i)$  and under the CSME statistical assumptions and notation,  $\hat{E}(Y | \mathbf{A} = \mathbf{a}, \mathbf{L}_i) = \beta_0 + \mathbf{a}\beta_a + \mathbf{l}_i\beta_l + \mathbf{a}\beta_{al}\mathbf{l}_i^T$ . Then the estimating function for  $\psi_k$  is unbiased, since

$$\begin{aligned} E_0(\beta_{a_k l} \mathbf{L}^T + \beta_{a_k} - \psi_k) &= E_0(\beta_{a_k l} \mathbf{L}^T + \beta_{a_k} - [E\{Y(a_k, \mathbf{a}_{(-k)})\} - E\{Y(a_k - 1, \mathbf{a}_{(-k)})\}]) \\ &= E_0 \left[ \beta_{a_k l} \mathbf{L}^T + \beta_{a_k} - \left\{ \frac{1}{n} \sum_i (\beta_0 + \mathbf{a}\beta_a + \mathbf{l}_i\beta_l + \mathbf{a}\beta_{al}\mathbf{l}_i^T) \right. \right. \\ &\quad \left. \left. - \frac{1}{n} \sum_i (\beta_0 + \mathbf{a}\beta_a - \beta_{a_k} + \mathbf{l}_i\beta_l + \mathbf{a}\beta_{al}\mathbf{l}_i^T - \beta_{a_k l} \mathbf{l}_i^T) \right\} \right] \\ &= E_0 \left\{ \beta_{a_k l} \mathbf{L}^T + \beta_{a_k} - \frac{1}{n} \sum_i (\beta_{a_k} + \beta_{a_k l} \mathbf{l}_i^T) \right\} \\ &= E_0 \left( \beta_{a_k l} \mathbf{L}^T - \frac{1}{n} \sum_i \beta_{a_k l} \mathbf{l}_i^T \right) \\ &= \beta_{a_k l} E_0(\mathbf{L}^T) - \beta_{a_k l} E_0(\mathbf{L}^T) \\ &= 0 \end{aligned}$$

where  $\mathbf{a} = (a_k, \mathbf{a}_{(-k)})$  are arbitrary constants and  $E_0$  refers to taking the expectation under the true parameter value.

Denote  $\hat{\Theta}$  as the solution to  $\sum_{i=1}^n \psi_{GF-CSME}(Y_i, \mathbf{L}_i, \mathbf{A}_i^*, \Sigma, \hat{\Theta}) = 0$ . Then by the proof above  $\sqrt{n}(\hat{\Theta} - \Theta) \sim N(\mathbf{0}, A^{-1}B(A^{-1})^T)$  where  $A$  and  $B$  are consistently estimated by

$$\hat{A} = \frac{1}{n} \sum_{i=1}^n \frac{d}{d\Theta^T} \psi_{GF-CSME}(Y_i, \mathbf{L}_i, \mathbf{A}_i^*, \Sigma, \hat{\Theta})$$

$$\hat{B} = \frac{1}{n} \sum_{i=1}^n \psi_{GF-CSME}(Y_i, \mathbf{L}_i, \mathbf{A}_i^*, \Sigma, \hat{\Theta}) \psi_{GF-CSME}^T(Y_i, \mathbf{L}_i, \mathbf{A}_i^*, \Sigma, \hat{\Theta})$$

In the R code implementing the methods, this sandwich variance estimation is accomplished using the R package `geex` (Saul and Hudgens, 2020).

### 1.1.2 Relationship to classical causal estimators

It has been noted (see Carroll et al. (2006)) that the CSME estimating equations reduce to the score equations for a GLM when the measurement error covariance matrix  $\Sigma = 0_{m \times m}$ . Thus under no measurement error, the procedure described above reduces to a stack of estimating equations corresponding to the common practice of performing the g-formula while specifying a GLM for the outcome regression, making it a special case of the proposed estimator.

## 1.2 IPW CSME estimator

### 1.2.1 Large sample properties

Consistency and asymptotic normality of the IPW CSME estimator are proven as above, using M-estimator theory. The partial M-estimator (Stefanski and Boos, 2002) corresponding to the parameters of interest is  $\sum_{i=1}^n \psi(Y_i, \mathbf{Z}_i, \mathbf{L}_i, \mathbf{A}_i^*, \Theta) = \sum_{i=1}^n SW_i(Y_i - E[Y_i | \Delta_i]) \Delta_i^T = 0$ . It suffices to show that the expectation of the estimating function  $\psi(Y_i, \mathbf{Z}_i, \mathbf{L}_i, \mathbf{A}_i^*, \Theta)$  is equal to 0. Although  $\Delta_i = (\Delta_{i1}, \dots, \Delta_{im})$  is a vector with length  $m$  for each individual  $i$ , the estimator form is the same for each row of the EE stack and without loss of generality, the estimating function is proven to be unbiased for  $SW_i(Y_i - E[Y_i | \Delta_i]) \Delta_{ik}$  for an arbitrary

$1 \leq k \leq m$ .

First consider a slightly different estimator, namely the estimator weighted by the unknown true propensity weights. Let  $SW = \frac{h(\mathbf{A})}{f(\mathbf{A}|\mathbf{L})}$  such that the numerator is any function of  $\mathbf{A}$  and the denominator is a conditional density of exposures given confounders. Furthermore, suppose that the denominator density equals the true conditional density, denoted  $f(\mathbf{A}|\mathbf{L}) = f_0(\mathbf{A}|\mathbf{L})$ . Let  $E_0$  notation refer to taking the expectation under the true causal parameter vector from the MSM, which is nested within  $E(Y|\Delta)$ . In a slight abuse of notation, let  $\Delta_k$  be the random variable corresponding to the  $k^{th}$  element of the random vector  $\Delta$ , rather than the vector  $\Delta$  for individual  $k$  which is instead notated as bold  $\Delta_k$  in the manuscript.

$$\begin{aligned}
E_0 \left[ \frac{h(\mathbf{A})}{f_0(\mathbf{A}|\mathbf{L})} \{Y - E(Y|\Delta)\} \Delta_k \right] &= E_0 \left( E \left[ \frac{h(\mathbf{A})}{f_0(\mathbf{A}|\mathbf{L})} \{Y^{\mathbf{A}} - E(Y^{\mathbf{A}}|\Delta)\} \Delta_k | \mathbf{L} \right] \right) \\
&= E_0 \left[ \int_{\mathbf{a}} \frac{h(\mathbf{a})}{f_0(\mathbf{a}|\mathbf{L})} \{Y^{\mathbf{a}} - E(Y^{\mathbf{a}}|\Delta)\} \Delta_k f_0(\mathbf{a}|\mathbf{L}) d\mu(\mathbf{a}) \right] \\
&= E_0 \left\{ \int_{\mathbf{a}} h(\mathbf{a}) (Y^{\mathbf{a}} - E[Y^{\mathbf{a}}|\Delta]) \Delta_k d\mu(\mathbf{a}) \right\} \\
&= \int_{\mathbf{a}} E_0 \{ h(\mathbf{a}) (Y^{\mathbf{a}} - E[Y^{\mathbf{a}}|\Delta]) \Delta_k \} d\mu(\mathbf{a}) \\
&= \int_{\mathbf{a}} E_0 \left\{ E \left[ h(\mathbf{a}) (Y^{\mathbf{a}} - E[Y^{\mathbf{a}}|\Delta]) \Delta_k | \Delta \right] \right\} d\mu(\mathbf{a}) \\
&= \int_{\mathbf{a}} E_0 \left\{ h(\mathbf{a}) \Delta_k E \left[ (Y^{\mathbf{a}} - E[Y^{\mathbf{a}}|\Delta]) | \Delta \right] \right\} d\mu(\mathbf{a}) \\
&= \int_{\mathbf{a}} E_0 \{ h(\mathbf{a}) \Delta_k (E[Y^{\mathbf{a}}|\Delta] - E[Y^{\mathbf{a}}|\Delta]) \} d\mu(\mathbf{a}) \\
&= 0
\end{aligned}$$

where  $d\mu(\mathbf{a})$  is defined as the Lebesgue measure. The first equality uses causal consistency, the second equality uses conditional exchangeability, and positivity is needed for the integral to be well-defined. Thus the estimator is consistent and asymptotically normal by standard M-estimator theory. The asymptotic variance is given by the usual sandwich estimator as

described in the previous section.

From here there are two jumps to the proposed estimator. The first is that even if no treatments were mismeasured, one would still need to estimate the treatment weights from some kind of model. This substitution is well known to result in a consistent estimator as long as the propensity score model is correctly specified. The second jump was alluded to in Section 3.3 of the paper, that the proposed estimator uses weights estimated from a propensity model that is fit using the mismeasured exposures. This will not necessarily affect the consistency of the estimator. For example, suppose the exposures are independent given  $\mathbf{L}$  and that each exposure has a linear relationship with the confounders, i.e., with simplified scalar notation,  $A = \mathbf{L}\alpha + \epsilon_{ps}$ . Then under additive measurement error, each mismeasured observed exposure also has a linear relationship with the confounders given by:  $A^* = \mathbf{L}\alpha + \epsilon_{ps} + \epsilon_{me}$ . So if linear propensity models are fit using the mismeasured exposures (noting that  $\mathbf{A}^*$  is a collider on the only path connecting  $\epsilon_{me}$  and  $\mathbf{L}$ ), one would still get consistent estimates of the propensity model parameters  $\alpha$  (and subsequently the weights), albeit with more variability. Therefore the proposed estimator would still be consistent. When exposures have non-linear, complex relationships with confounders, consistency may not be guaranteed, but the measurement error will likely only introduce mild issues (Carroll et al., 2006).

### 1.2.2 Relationship to classical causal estimators

Note that when the measurement error covariance matrix  $\Sigma = \mathbf{0}_{m \times m}$ , the sufficient statistic  $\Delta$  reduces to the observed exposure vector. Then the IPW estimator reduces to the form  $\sum_{i=1}^n \psi(Y_i, \mathbf{L}_i, \mathbf{A}_i^*, \Theta) = 0$  where:

$$\psi(Y_i, \mathbf{L}_i, \mathbf{A}_i^*, \Theta) = \begin{bmatrix} SW_i \{Y_i - E(Y_i | \mathbf{A}_i^*)\} (1, \mathbf{A}_i^*)^T \\ SW_i \left[ \phi - \frac{\{Y_i - E(Y_i | \mathbf{A}_i^*)\}^2}{Var(Y_i | \mathbf{A}_i^*) / \phi} \right] \end{bmatrix} \quad (1)$$

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 CSME estimator where there is no measurement error present.

## 1.3 Doubly Robust Estimator

### 1.3.1 Large Sample Properties

Once again, consistency and asymptotic normality of the proposed estimator is proven using M-estimator theory. The estimating function of the DR estimator has the form

$$\frac{h(\mathbf{A})}{f(\mathbf{A}|\mathbf{L})}[\{Y - E(Y|\Delta)\}\Delta - Q(\mathbf{A}, \mathbf{L})] + \int_{\mathbf{a}} h(\mathbf{a})Q(\mathbf{A} = \mathbf{a}, \mathbf{L})d\mu(\mathbf{a}) \quad (2)$$

where  $f(\mathbf{A}|\mathbf{L})$  and  $h(\mathbf{A})$  are defined as before and  $Q(\mathbf{A}, \mathbf{L})$  is a function of  $\mathbf{A}$  and  $\mathbf{L}$  that often represents the outcome regression. One must show that the expected value of the estimating function at the true parameter value  $\gamma_0$  is equal to 0 if either the propensity model or outcome regression model is correctly specified.

First, suppose that the outcome regression model is correctly specified, denoted  $Q(\mathbf{A}, \mathbf{L}) = Q_0(\mathbf{A}, \mathbf{L}) = E[Y|A, L] - \gamma_0(1, \mathbf{a})^T$ . Let  $E_0$  denote taking the expectation under true parameter  $\gamma_0$ , then

$$\begin{aligned} & E \left[ \frac{h(A)}{f(A|L)} [\{Y - E(Y|L, \Delta)\}\Delta - Q_0(\Delta, L)] + \int_a h(a)Q_0(A = a, L)d\mu(a) \right] \\ &= E \left[ \frac{h(A)}{f(A|L)} [\{Y - E(Y|L, \Delta)\}\Delta - Q_0(A, L)] \right] + E \left[ \int_a h(a)E[\epsilon(\beta_0)|A = a, L]d\mu(a) \right] \\ &= E \left\{ E \left[ \frac{h(A)}{g(A|L)} [\epsilon(\beta_0) - Q_0(A, L)] | A, L \right] \right\} + E \left\{ E \left[ \int_a h(a)E[\epsilon(\beta_0)|A = a, L]d\mu(a) | V \right] \right\} \\ &= E \left\{ \frac{h(A)}{g(A|L)} [Q_0(A, L) - Q_0(A, L)] \right\} + E \left\{ \int_a h(a)E[\epsilon(\beta_0)|A = a, V]d\mu(a) \right\} \\ &= E \left\{ \int_a h(a)E[\epsilon(\beta_0)|A = a, V]d\mu(a) \right\} \end{aligned}$$

$$\begin{aligned}
&= E \left\{ \int_a h(a) E[Y^A - E[Y^A|V]|A = a, V] d\mu(a) \right\} \\
&= E \left\{ \int_a h(a) (E[Y^a|V] - E[Y^a|V]) d\mu(a) \right\} \\
&= 0
\end{aligned}$$

$$\begin{aligned}
&E \left[ \frac{h(A)}{f(A|L)} [\epsilon(\beta_0) - Q_0(A, L)] + \int_a h(a) Q_0(A = a, L) d\mu(a) \right] \\
&= E \left[ \frac{h(A)}{f(A|L)} [\epsilon(\beta_0) - Q_0(A, L)] \right] + E \left[ \int_a h(a) E[\epsilon(\beta_0)|A = a, L] d\mu(a) \right] \\
&= E \left\{ E \left[ \frac{h(A)}{g(A|L)} [\epsilon(\beta_0) - Q_0(A, L)] | A, L \right] \right\} + E \left\{ E \left[ \int_a h(a) E[\epsilon(\beta_0)|A = a, L] d\mu(a) | V \right] \right\} \\
&= E \left\{ \frac{h(A)}{g(A|L)} [Q_0(A, L) - Q_0(A, L)] \right\} + E \left\{ \int_a h(a) E[\epsilon(\beta_0)|A = a, V] d\mu(a) \right\} \\
&= E \left\{ \int_a h(a) E[\epsilon(\beta_0)|A = a, V] d\mu(a) \right\} \\
&= E \left\{ \int_a h(a) E[Y^A - E[Y^A|V]|A = a, V] d\mu(a) \right\} \\
&= E \left\{ \int_a h(a) (E[Y^a|V] - E[Y^a|V]) d\mu(a) \right\} \\
&= 0
\end{aligned}$$

Now suppose the outcome regression is potentially misspecified. We show the estimator is unbiased as long as the propensity score model is correctly specified, denoted  $g(A|L) = g_0(A|L)$ .

$$\begin{aligned}
&E \left[ \frac{h(A)}{f_0(A|L)} \left[ \{Y - E(Y|\Delta)\} \Delta_k - Q(A, L) \right] + \int_a h(a) Q(A = a, L) d\mu(a) \right] \\
&= E \left[ \frac{h(A)}{f_0(A|L)} \{Y - E(Y|\Delta)\} \Delta_k \right] - E \left[ \frac{h(A)}{f_0(A|L)} Q(A, L) \right] + E \left[ \int_a h(a) Q(A = a, L) d\mu(a) \right]
\end{aligned}$$

Since the left term is an unbiased estimating function for  $\beta$  as shown earlier, it suffices

to show that

$$\int_a E \left[ h(a) Q(A = a, L) \right] d\mu(a) - E \left[ \frac{h(A)}{f_0(A|L)} Q(A, L) \right] = 0$$

To see this, note that

$$\begin{aligned} E \left[ \frac{h(A)}{g_0(A|L)} Q(A, L) \right] &= E \left\{ E \left[ \frac{h(A)}{g_0(A|L)} Q(A, L) | L \right] \right\} \\ &= E \left\{ \int_a \frac{h(a)}{g_0(a|L)} Q(A = a, L) g_0(a|L) d\mu(a) \right\} \\ &= \int_a E \left[ h(a) Q(A = a, L) \right] d\mu(a) \end{aligned}$$

### 1.3.2 Relationship to classical causal estimators

Insert description, under no measurement error it reduces to the DR estimator from Neugebauer and van der Laan (2005).

## 1.4 Uniqueness of EE solutions

Each of the proofs above relies on there being a unique solution to each set of estimating equations. It has been noted in prior work (?) that similar conditional score equations do not always have unique solutions, but that multiple solutions are very rare in practice. In the various simulations of this paper, multiple solutions 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 watch for rare instances of multiple solutions, unusual estimates, and/or root-solving algorithm divergence errors.



## 2 Web Appendix B: Accounting for Two-Phase Sampling

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, we demonstrate one such approach under a using a simulation study. In particular, for this simulation and the application section analysis, we follow the simple IPW method described in Wang et al. (2009), but the DR approaches from the same paper or from Rose and van der Laan (2011) could also be explored to account for two-phase sampling within our proposed methods. The simple IPW method is implemented by weighting each of the proposed estimating equations by the inverse probability of selection for the second-phase of the study (multiplying treatment weights by sampling weights for the IPW-CSME and DR-CSME estimators) and restricting the analysis to those selected. This works well for the subset of the HVTN 505 trial that is the focus of Section 2.5, but may be inadequate when analyzing exposures measured in a subsample in conjunction with exposures measured in the full sample.

### 2.1 Two-phase sampling simulations

We replicate the structure of the second simulation study described in Section 5 of the paper, but under a two-phase sampling design. \*\*\*These are a few months old and might not be fully up to date with most recent versions of main sims, double-check before adding here\*\*\*. The results are presented in the following table:

### 3 Web Appendix C: Additional Simulations

In this section, we evaluate the performance of the methods under two assumption violations: (i) when positivity doesn't hold and (ii) when measurement error doesn't follow a classical additive model.

#### 3.1 Under positivity violation

To evaluate the proposed methods under positivity violations we replicate the general structure of the first simulation study from Section 5 of the paper almost exactly. We create a strong positivity violation (lack of positivity in reasonably sized finite samples rather than a structural lack of positivity) by changing how the treatment  $A_1$  is generated from  $N(4+L_1, 1)$  to  $N(4+L_1, 0.2)$ . This breaks the phenomenon of mostly overlapping treatment values experienced by simulated subjects with  $L_1 = 1$  and  $L_1 = 0$ , 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 the following table.

Table 1. *Simulation study under positivity violation*

| Estimator      | $\psi_1$ |       |       |          | $\psi_2$ |       |       |          | $\psi_3$ |       |       |          |
|----------------|----------|-------|-------|----------|----------|-------|-------|----------|----------|-------|-------|----------|
|                | Bias     | ASE   | ESE   | Coverage | Bias     | ASE   | ESE   | Coverage | Bias     | ASE   | ESE   | Coverage |
| Regression     | -0.008   | 0.022 | 0.022 | 93.6%    | 0.070    | 0.044 | 0.045 | 69.4%    | 0.010    | 0.023 | 0.023 | 91.8%    |
| CSME           | 0.008    | 0.024 | 0.024 | 93.4%    | 0.070    | 0.044 | 0.045 | 68.2%    | -0.016   | 0.038 | 0.038 | 93.4%    |
| G-formula      | -0.068   | 0.016 | 0.015 | 0.8%     | 0.000    | 0.037 | 0.037 | 94.0%    | 0.010    | 0.023 | 0.023 | 91.2%    |
| IPW            | -0.068   | 0.022 | 0.024 | 19.8%    | -0.003   | 0.057 | 0.061 | 96.2%    | 0.010    | 0.033 | 0.033 | 92.8%    |
| G-formula CSME | -0.057   | 0.017 | 0.016 | 7.2%     | 0.000    | 0.038 | 0.037 | 93.8%    | -0.016   | 0.038 | 0.038 | 93.4%    |
| IPW CSME       | -0.058   | 0.024 | 0.026 | 35.2%    | -0.004   | 0.057 | 0.060 | 96.2%    | -0.017   | 0.054 | 0.055 | 94.4%    |

We note that the results overall look similar to that in Table 1 of the paper. However, with the positivity assumption broken, the proposed methods do not perform as well in estimating the effect of  $A_1$ . In addition, for the proposed weighted-CSME estimator, some bias seems to bleed over into the estimation of effects of the other treatments for which positivity still holds. However, the proposed methods still greatly outperform all comparator methods in this scenario.

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".

### 3.2 Under non-additive measurement error

(Can probably remove this section, I like to see simulations under assumption violations but these didn't turn out particularly interesting)

Next we evaluate the proposed methods when treatment measurement error does not follow the classical additive model. In particular, we once again replicate the first simulation study from Section 5 of the paper, but change the simulation of mismeasured treatments  $A_1$  and  $A_3$  such that  $A_1$  now follows a multiplicative error model and  $A_3$  follows an additive model where the magnitude of error depends on an unobserved variable  $C$ . In particular, we simulate  $A_1^* = A_1 \epsilon_{me1}$  where  $\epsilon_{me1} \sim N(1, 0.15)$  and  $A_3^* = A_3 + \epsilon_{me3}$  where  $\epsilon_{me3} \sim N(0, 0.2 + 0.3C)$  and  $C \sim \text{Binom}(0.5)$ . The methods are still performed assuming additive measurement error with known measurement error covariance as specified in section 5 of the paper. The results are presented in the following table.

Table 2. *Simulation study under non-additive measurement error*

| Estimator      | $\psi_1$ |       |       |          | $\psi_2$ |       |       |          | $\psi_3$ |       |       |          |
|----------------|----------|-------|-------|----------|----------|-------|-------|----------|----------|-------|-------|----------|
|                | Bias     | ASE   | ESE   | Coverage | Bias     | ASE   | ESE   | Coverage | Bias     | ASE   | ESE   | Coverage |
| Regression     | 0.032    | 0.023 | 0.024 | 75.4%    | 0.070    | 0.043 | 0.045 | 68.0%    | 0.013    | 0.022 | 0.022 | 88.8%    |
| CSME           | 0.053    | 0.025 | 0.026 | 46.6%    | 0.070    | 0.044 | 0.044 | 67.6%    | -0.008   | 0.035 | 0.034 | 95.0%    |
| G-formula      | -0.036   | 0.017 | 0.017 | 39.4%    | 0.000    | 0.037 | 0.036 | 94.2%    | 0.013    | 0.022 | 0.022 | 88.6%    |
| IPW            | -0.036   | 0.025 | 0.028 | 77.4%    | -0.003   | 0.058 | 0.061 | 95.6%    | 0.013    | 0.032 | 0.032 | 92.6%    |
| G-formula CSME | -0.021   | 0.018 | 0.018 | 77.4%    | 0.000    | 0.037 | 0.036 | 94.4%    | -0.008   | 0.035 | 0.034 | 95.0%    |
| IPW CSME       | -0.021   | 0.027 | 0.030 | 91.0%    | -0.003   | 0.058 | 0.061 | 95.6%    | -0.008   | 0.050 | 0.051 | 94.8%    |

The proposed methods still seem to perform well for treatment  $A_3$ . For treatment  $A_1$  there is some finite-sample bias, but the proposed methods continued to outperform the

comparator methods.

## 4 Web Appendix D: More complex model specifications

The three proposed methods are based on a linear marginal structural model form. While this is helpful to match the conditional score framework described in Section 3.1, it is too restrictive for some potential applications. To this end we note that transformations of elements of  $\mathbf{A}$  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 a transformation of an exposure is assumed to follow a multiplicative measurement error model then that variable cannot be included in the MSM. However, if the variable is strictly positive, then its log transform would 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.

Any other options for if additive ME doesn't hold?

## References

- Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement error in nonlinear models: a modern perspective*. CRC press.
- 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.
- Rose, S. and van der Laan, M. J. (2011). A targeted maximum likelihood estimator for two-stage designs. *The International Journal of Biostatistics* **7**,.

- Saul, B. and Hudgens, M. (2020). The calculus of m-estimation in r with geex. *Journal of Statistical Software, Articles* **92**, 1–15.
- Stefanski, L. A. and Boos, D. D. (2002). The calculus of m-estimation. *The American Statistician* **56**, 29–38.
- 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.