Web Appendix for "Addressing Confounding and Exposure Measurement Error Using Conditional Score Functions"

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. Thus, the proposed g-formula CSME estimator can be written as $\sum_{i=1}^{n} \psi_{GF-CSME}(Y_i, \boldsymbol{L}_i, \boldsymbol{A}_i^*, \Sigma_{me}, \Theta_{GF}) = 0$, where $\Theta_{GF} = (\beta_0, \beta_a^T, \beta_l^T, vec(\beta_{al}), \phi, E\{Y(\boldsymbol{a})\})$ and:

$$\psi_{GF-CSME}(Y, \boldsymbol{L}, \boldsymbol{A}^*, \Sigma_{me}, \Theta_{GF}) = \begin{bmatrix} \{Y - \mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta})\}(1, \boldsymbol{L}, \boldsymbol{\Delta}, \boldsymbol{L} \otimes \boldsymbol{\Delta})^T \\ \phi - \{Y - \mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta})\}^2 / \{\mathrm{Var}(Y|\boldsymbol{L}, \boldsymbol{\Delta})/\phi\} \\ g^{-1}(\beta_0 + \boldsymbol{a}\beta_a + \boldsymbol{l}\beta_l + \boldsymbol{a}\beta_{al}\boldsymbol{l}^T) - E\{Y(\boldsymbol{a})\} \end{bmatrix}$$

The parameter of interest $E\{Y(\boldsymbol{a})\}$ is in the last estimating equation of the stack, and the proof relies on (i) the usual g-formula proof based on the causal assumptions made in Section 3.1 of the paper and (ii) that the CSME estimator $\hat{E}(Y|\boldsymbol{A}=\boldsymbol{a},\boldsymbol{L})$ was previously shown to be consistent (Carroll et al., 2006):

$$\begin{aligned}
E[g^{-1}(\beta_0+\boldsymbol{a}\beta_a+\boldsymbol{l}\beta_l+\boldsymbol{a}\beta_{al}\boldsymbol{l}^T)-E\{Y(\boldsymbol{a})\}] &= E\{g^{-1}(\beta_0+\boldsymbol{a}\beta_a+\boldsymbol{l}\beta_l+\boldsymbol{a}\beta_{al}\boldsymbol{l}^T)\}-E\{Y(\boldsymbol{a})\} \\
&= E\{\hat{E}(Y|\boldsymbol{A}=\boldsymbol{a},\boldsymbol{L})\}-E[E\{Y(\boldsymbol{a})|\boldsymbol{L}\}] \\
&= E\{E(Y|\boldsymbol{A}=\boldsymbol{a},\boldsymbol{L})\}-E[E\{Y(\boldsymbol{a})|\boldsymbol{A}=\boldsymbol{a},\boldsymbol{L}\}] \\
&= E\{E(Y|\boldsymbol{A}=\boldsymbol{a},\boldsymbol{L})\}-E\{E(Y|\boldsymbol{A}=\boldsymbol{a},\boldsymbol{L})\} \\
&= 0
\end{aligned}$$

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

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

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

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 $\Sigma_{me}=0_{m\times m}$. 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 $\sum_{i=1}^{n} \psi(Y_i, \boldsymbol{Z}_i, \boldsymbol{L}_i, \boldsymbol{A}_i^*, \Theta_{IPW}) = \sum_{i=1}^{n} SW_i(Y_i - E[Y_i|\boldsymbol{\Delta}_i])\boldsymbol{\Delta}_i^T = 0$. It suffices to show that the expectation of the estimating function $\psi(Y, \boldsymbol{Z}, \boldsymbol{L}, \boldsymbol{A}^*, \Theta_{IPW})$ is equal to 0. Although $\boldsymbol{\Delta} = (\Delta_1, ..., \Delta_m)$ is a vector of length m, the estimator form is the same for each row of the estimating equation stack and without loss of generality, the estimating function is proven to be unbiased for $SW(Y - E[Y|\boldsymbol{\Delta}])\Delta_k$ for an arbitrary $1 \leq k \leq m$.

First consider the estimator as written, where it is weighted by the true propensity weights which aren't typically known in practice. Let $SW = \frac{h(\boldsymbol{A})}{f(\boldsymbol{A}|\boldsymbol{L})}$ such that the numerator is any function of $\boldsymbol{A}$ and the denominator is a conditional density of exposures given confounders. Furthermore, suppose that the denominator density equals the true conditional density, denoted $f(\boldsymbol{A}|\boldsymbol{L}) = f_0(\boldsymbol{A}|\boldsymbol{L})$. Let E_0 notation refer to taking the expectation under the true causal parameter vector from the MSM, which is nested within $E(Y|\boldsymbol{\Delta})$. In a slight abuse of notation, let Δ_k be the random variable corresponding to the k^{th} element of the random vector $\boldsymbol{\Delta}$, rather than the vector $\boldsymbol{\Delta}$ for individual k which is instead notated as bold $\boldsymbol{\Delta}_k$ in the manuscript.

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

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

where $d\mu(\boldsymbol{a})$ 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 corresponding estimator when weights are estimated. The first is that one needs to estimate the treatment weights from some kind of model, even if no treatments were mismeasured. This substitution is well known to result in a consistent estimator as long as the propensity score model is correctly specified, because then this estimator will equal the estimator described above plus an $o_p(1)$ term. The second jump was alluded to in Section 3.3 of the paper, that one can use 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 $\boldsymbol{L}$ and that each exposure has a linear relationship with the confounders, i.e., with simplified scalar notation, $A = \boldsymbol{L}\alpha + \epsilon_{ps}$. Then under additive measurement error, each mismeasured observed exposure also has a linear relationship with the confounders given by: $A^* = \boldsymbol{L}\alpha + \epsilon_{ps} + \epsilon_{me}$. So if linear propensity models are fit using the mismeasured exposures (noting that $\boldsymbol{A}^*$ is a collider on the only path connecting ϵ_{me} and $\boldsymbol{L}$), one would still get consistent estimates of the propensity model parameters α (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 previous explorations of this topic suggest that 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 $\Sigma_{me} = \mathbf{0}_{m\times m}$, 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:

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

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 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. This estimator is a solution to the estimating equation $\sum_{i=1}^{n} \psi_{DR-CSME}(Y_i, \boldsymbol{L}_i, \boldsymbol{A}_i^*, \Sigma_{me}$ 0, where $\Theta_{DR} = \Theta_{GF}$ and

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

First, suppose that the outcome regression is correctly specified. Need to take expectation and show weighted CSME is equivalent to CSME in expectation...?

Then suppose that the propensity model is correctly specified.

The estimating function of the DR estimator has the form

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

where $f(\boldsymbol{A}|\boldsymbol{L})$ and $h(\boldsymbol{A})$ are defined as before and $Q(\boldsymbol{A}, \boldsymbol{L})$ is a function of $\boldsymbol{A}$ and $\boldsymbol{L}$ that often represents the outcome regression. One must show that the expected value of the estimating function at the true parameter value γ_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 $Q(\boldsymbol{A}, \boldsymbol{L}) = Q_0(\boldsymbol{A}, \boldsymbol{L}) = E[Y|A, L] - \gamma_0(1, \boldsymbol{a})^T$. Let E_0 denote taking the expectation under true parameter γ_0, then

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

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

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 $g(A|L) = g_0(A|L)$.

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

Since the left term is an unbiased estimating function for β 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

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

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 (Stefanski and Carroll, 1987) 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 2000 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, one such approach is demonstrated using a simulation study. In particular, for this simulation and the application section analysis, the simple IPW method described in Wang et al. (2009) is implemented, 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 the 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 5 of the paper, but may be inadequate when analyzing exposures measured in a sub-sample in conjunction with exposures measured in the full sample.

2.1 Two-phase sampling simulations

The structure of the first simulation study described in Section 4 of the 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 Appendix Figure 1 and Appendix Table 1, with Bias, ASE, ESE, and Coverage defined as in Table 1 of the paper.

The methods seem to perform well as they did in the full-sampling simulation provided in the 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 CSME estimator with sampling weights converged in all analyses presented in Section 5 of the paper.

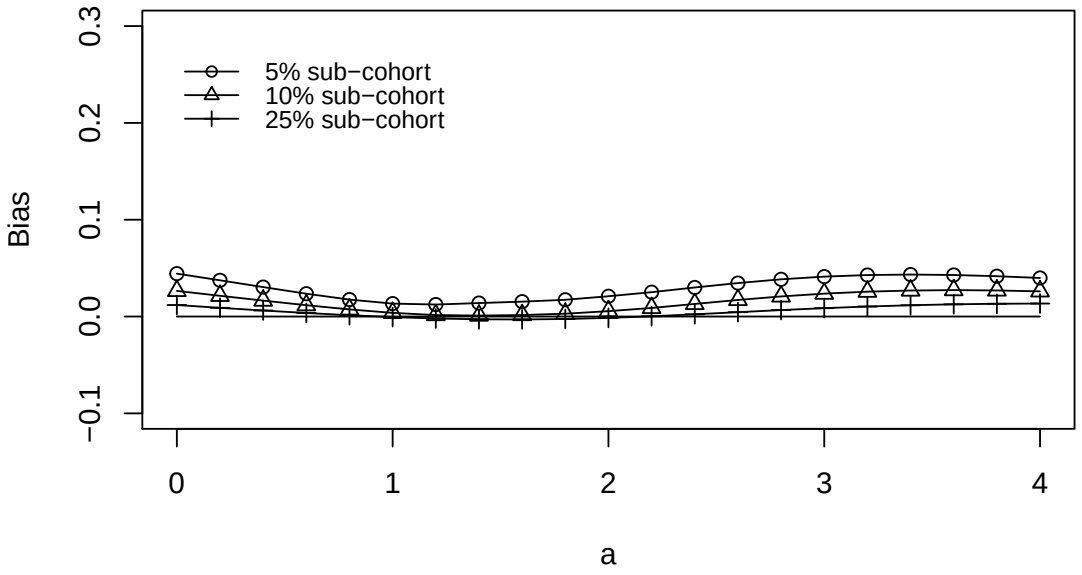

Figure 1: Estimated dose-response curve bias for the DR CSME method compared across three sub-cohort sizes. Bias refers to the average bias across 2,000 simulated data sets for each method evaluated at each point on the horizontal axis $a = (0, 0.2, 0.4, ..., 4)$.

Appendix Table 1. *Simulation study for case-cohort design*

Sub-cohort Size	Bias	ASE	ESE	Coverage	Percent failed to converge
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%

3 Web Appendix C: Additional Simulations

In this section, the methods are studied 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, the general structure of the first simulation study from Section 4 of the paper is replicated almost exactly. A strong positivity violation is created by changing how the treatment A_1 is generated from $\mathcal{N}(4 + L_1, 1)$ to $\mathcal{N}(4+L_1, 0.2)$. 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 Appendix Table 3.

Appendix Table 2. *Simulation study under positivity violation*

	ψ_1				ψ_2				ψ_3			
Estimator	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%

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 generally perform okay versus the 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

Next the proposed methods are evaluated when treatment measurement error does not follow the classical additive model. In particular, the first simulation study from Section 5 of the paper is again replicated, but the simulation of mismeasured treatments A_1 and A_3 are changed such that A_1 now follows a multiplicative error model and A_3 follows on additive model where the magnitude of error depends on an unobserved variable C. In particular, they were simulated as $A_1^* = A_1\epsilon_{me_1}$ where $\epsilon_{me_1} \sim \mathcal{N}(1, 0.15)$ and $A_3^* = A_3 + \epsilon_{me_3}$ where $\epsilon_{me_3} \sim \mathcal{N}(0, 0.2 + 0.3C)$ 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 Appendix Table 4. 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.

Appendix Table 3. *Simulation study under non-additive measurement error*

	ψ_1				ψ_2				ψ_3			
Estimator	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%

4 Web Appendix D: More complex model specifications for the IPW CSME estimator

The proposed IPW CSME estimator assumes 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 $\boldsymbol{A}$ 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 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.

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.

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.

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.

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.

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.