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.

Note-Should be no bold in beta subscripts - fix later.

1.1 G-formula 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). In particular, we begin by rewriting the estimator into the form of an M-estimator. Note that the CSME estimator is an M-estimator and that the g-formula can be written in the form of an unbiased estimating equation (EE). So we can write the g-formula CSME estimator as $\sum_{i=1}^{n} \psi_{GF-CSME}(Y_i, \mathbf{L}_i, \mathbf{A}^*_i, \Sigma, \Theta) = 0$, where $\Theta = (\beta_0, \beta_a^T, \beta_l^T, \phi, \psi_1, ..., \psi_m)$ and:

$$\psi_{GF-CSME}(Y_i, \mathbf{L}_i, \mathbf{A}^*{}_i, \Sigma, \Theta) = \begin{bmatrix} (Y_i - E[Y_i|\mathbf{L}_i, \Delta_i]) \begin{pmatrix} 1 \\ \mathbf{L}_i \\ \Delta_i \end{pmatrix} \\ \phi - \frac{(Y_i - E[Y_i|\mathbf{L}_i, \Delta_i]))^2}{Var[Y_i|\mathbf{L}_i, \Delta_i]/\phi} \\ \beta_{a_1\mathbf{l}}\mathbf{l}_i + \beta_{a_1} - \psi_1 \\ \vdots \\ \beta_{a_m\mathbf{l}}\mathbf{l}_i + \beta_{a_m} - \psi_m \end{bmatrix} \tag{1}$$

where notation for the coefficients from the CSME method are expanded as $\beta_{\mathbf{a}} = (\beta_{a_1}, ..., \beta_{a_m})^T$ and $\beta_{\mathbf{al}} = (\beta_{a_1\mathbf{l}}, ..., \beta_{a_m\mathbf{l}})^T$, where $\beta_{a_j\mathbf{l}}$ is the vector of all interactions between a_j and elements of $\mathbf{L}$ for $j = 1, ..., m$ (with zeroes for interactions not included in the outcome model).

Each of the final m estimating equations in the stack is an algebraic simplification of the g-formula in this context, and the EE for an arbitrary parameter ψ_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, $E[Y(\mathbf{a})] = \frac{1}{n}\sum_i \hat{E}[Y|\mathbf{A} = \mathbf{a}, \mathbf{L}_i]$ and under the CSME statistical assumptions and notation, $\hat{E}[Y|\mathbf{A} = \mathbf{a}, \mathbf{L}_i] = \beta_0 + \beta_a^T \mathbf{a} + \beta_\mathbf{l}^T \mathbf{l}_i + \beta_{\mathbf{al}}^T (a \otimes \mathbf{l}_i)$. Then the estimating function for ψ_k is unbiased, since

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

$$= 0$$

where $\mathbf{a} = (a_k, \mathbf{a}_{(-k)})$ are arbitrary values in the support of the exposure space 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 $\Sigma = 0$. 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. This is a special case of our proposed estimator.

1.2   IPW estimator

1.2.1   Large sample properties

We prove consistency and asymptotic normality as above, using M-estimator theory. Note that the partial M-estimator (Stefanski and Boos, 2002) corresponding to the parameters of interest is $\sum_{i=1}^{n} \psi(Y_i, \mathbf{Z}_i, \mathbf{L}_i, \mathbf{A}^*_i, \Theta) = \sum_{i=1}^{n} SW_i(Y_i - E[Y_i|\Delta_i])\Delta_i = 0$. 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

Δ_i is a vector with length m for each individual i, the estimator form is the same for each component and without loss of generality, we prove that the estimating function is unbiased for scalar Δ_i.

We begin by proving a slightly different estimator is consistent and asymptotically normal, namely the estimator weighted by the unknown true propensity weights. Let $SW = \frac{h(A)}{f(A|L)}$ such that the numerator is any function of A and the denominator is a conditional density of exposures given confounders. Furthermore, suppose that the denominator density equals the true conditional density, denoted $f(A|L) = f_0(A|L)$. Let E_0 notation refer to taking the expectation under the true causal parameter vector ψ_0 (which is nested within $E[Y|\Delta]$).

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

where $d\mu(a)$ 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 our proposed estimator. The first is that even if no treatments were mismeasured, we'd still need to estimate the treatment weights from some kind of model. This substitution is well known to result in an unbiased estimator as long as the propensity score model is correctly specified. The second jump was described in Section 3.3.3 of the paper, that the proposed estimator uses weights estimated from a propensity model that is fitting using the mismeasured treatments. Under the assumptions described in that section, the estimated weights are consistent when the propensity model is correctly specified and thus the proof above holds for the proposed estimator as well.

1.2.2 Relationship to classical causal estimators

Note that when the measurement error covariance matrix $\Sigma = \mathbf{0}$, the sufficient statistic Δ reduces to the observed treatment 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:

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

This is exactly the score function vector for a GLM weighted by SW. Thus, the IPW estimator fit using a weighted GLM for outcome Y is a special case of our proposed estimator where there is no measurement error. Our estimator can be thought of as an extension of the traditional IPW estimator which allows for measurement error, but restricted to the GLM setting.

1.3 Doubly Robust Estimator

Still changing notation over from old version of proof, don't bother reading this section until I get rid of the ϵ notation that isn't helpful here.

1.3.1 Large Sample Properties

We once again prove consistency and asymptotic normality of the proposed estimator using M-estimator theory. The estimating function of the DR estimator has the form

$$\frac{h(A)}{f(A|L)}[Y - E[Y|\Delta] - Q(A, L)]\Delta + \int_a h(a, V)Q(A = a, L)d\mu(a) \tag{3}$$

where $f(A|L)$ and $h(A)$ are defined as before and $Q(A, L)$ is a function of A and L that often represents the outcome regression. We must show that the expected value of the estimating function at the true parameter value ψ_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 $Q(A, L) = Q_0(A, L) = E[Y|A, L] - \psi_0^T(1, \mathbf{a})^T$. Let E_0 denote taking the expectation under true parameter ψ_0, then

$$E\left[\frac{h(A)}{f(A|L)}[\epsilon(\beta_0) - Q_0(A, L)] + \int_a h(a)Q_0(A = a, L)d\mu(a)\right]$$

$$= E\left[\frac{h(A)}{f(A|L)}[\epsilon(\beta_0) - Q_0(A, L)]\right] + E\left[\int_a h(a)E[\epsilon(\beta_0)|A = a, L]d\mu(a)\right]$$

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

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

$$= E\left\{\int_a h(a)E[\epsilon(\beta_0)|A = a, V]d\mu(a)\right\}$$

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

$$= E\left\{\int_a h(a)(E[Y^a|V] - E[Y^a|V])d\mu(a)\right\}$$

$$= 0$$

Now suppose the outcome regression is potentially misspecified. We show the estimator

6

is unbiased as long as the propensity score model is correctly specified, denoted $g(A|L) = g_0(A|L)$.

$$E\left[\frac{h(A,V)}{g_0(A|L)}[\epsilon(\beta_0) - Q(A,L)] + \int_a h(a,V)Q(A=a,L)d\mu(a)\right]$$
$$= E\left[\frac{h(A,V)}{g_0(A|L)}\epsilon(\beta_0)\right] - E\left[\frac{h(A,V)}{g_0(A|L)}Q(A,L)\right] + E\left[\int_a h(a,V)Q(A=a,L)d\mu(a)\right]$$

Since the left term is an unbiased estimating function for β as shown earlier, it suffices to show that

$$\int_a E\left[h(a,V)Q(A=a,L)\right]d\mu(a) - E\left[\frac{h(A,V)}{g_0(A|L)}Q(A,L)\right] = 0$$

To see this, note that

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

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 Additional considerations

Might want to add a short discussion on (i) using additional variables in the numerator model and (ii) accounting for estimation of Σ or treatment weights or sampling weights when calculating the sandwich variance estimator. These are all covered in previous work so

don't need to provide much detail, maybe can just mention briefly and provide references?

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 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*

| | 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%

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 on 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 $C \sim Binom(0.5)$. 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	ψ_1 Bias	ASE	ESE	Coverage	ψ_2 Bias	ASE	ESE	Coverage	ψ_3 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 many 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.

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.