Biometrics 000, 000–000 DOI: 000
000 0000

Addressing Confounding and Exposure Measurement Error Using Conditional Score Functions

Bryan S. Blette[1], Peter B. Gilbert[2], and Michael G. Hudgens[1,*]

[1]Department of Biostatistics, University of North Carolina at Chapel Hill

Chapel Hill, NC, U.S.A.

[2]Department of Biostatistics, University of Washington and Fred Hutchinson Cancer Research Center

Seattle, Washington, U.S.A.

email: mhudgens@email.unc.edu

Summary: Confounding and measurement error are common barriers to drawing causal inference. While there are broad methodologies for addressing each phenomenon individually, confounding and measurement biases frequently co-occur and there is a paucity of methods that address them simultaneously. The few existing methods that do so rely on supplemental data or strong distributional and extrapolation assumptions to correct for measurement error. In this paper, methods are derived which instead leverage the likelihood structure under classical additive measurement error to draw inference using only measured variables. Three estimators are proposed based on g-computation, inverse probability weighting, and doubly-robust estimation techniques. The estimators are shown to be consistent and asymptotically normal, and the doubly-robust estimator is shown to exhibit its namesake property. The methods perform well in finite samples under both confounding and measurement error as demonstrated by simulation studies. The proposed doubly-robust estimator is applied to study the effects of two biomarkers on HIV-1 infection using data from the HVTN 505 vaccine trial.

Key words: Causal inference; Confounding; G-formula; HIV/AIDS; Marginal structural models; Measurement error.

This paper has been submitted for consideration for publication in *Biometrics*

1. Introduction

Confounding bias and measurement error are common barriers to identification, estimation, and inference of causal effects. These phenomena often occur together, but are rarely both addressed in an analysis, with researchers typically focusing on whichever is more egregious in their study or ignoring both entirely. The last few decades witnessed a proliferation of interest in and development of methods for causal inference and a parallel trend for methods accounting for measurement error, but comparatively few methods exist at the important intersection of these fields.

The measurement error literature is commonly split into (i) functional methods, which make no or limited assumptions on the distribution of mismeasured variables, and (ii) structural methods, which make explicit distributional assumptions (Carroll et al., 2006). Several causal methods based on structural approaches have been developed (Kuroki and Pearl, 2014; Edwards et al., 2015; Braun et al., 2017; Hong, Rudolph, and Stuart, 2017). Likewise, three of the four most popular functional methods (regression calibration, SIMEX, and methods based on instrumental variables) have been adapted to a variety of causal problems (Vansteelandt, Babanezhad, and Goetghebeur, 2009; Cole et al., 2010; Kendall, 2015; Lockwood and McCaffrey, 2015; Kyle et al., 2016; Wu et al., 2019). These methods all rely on either supplemental data, such as replication, validation, or instrumental data, or on strong distributional and extrapolation assumptions to draw inference.

In contrast, the fourth main functional approach, that of score functions, leverages the likelihood structure under classical additive measurement error to draw inference without supplemental data or strong assumptions. Despite this advantage, this method has only been adapted to perform causal inference in very limited capacity. In particular, McCaffrey, Lockwood, and Setodji (2013) suggested using weighted conditional score equations to correct for measurement error in confounders and the approach was eventually implemented in Shu and

Yi (2019). In this paper, this methodology is expanded in multiple directions by considering exposure/treatment measurement error and defining g-formula, inverse probability weighted (IPW), and doubly-robust (DR) estimators.

The proposed methods are motivated by the HVTN 505 vaccine trial. This HIV vaccine efficacy trial was stopped early after reaching predetermined cutoffs for futility (Hammer et al., 2013). However, subsequent analyses of trial data identified several correlates of risk among trial participants (Decamp et al., 2017; Janes et al., 2017; Fong et al., 2018; Neidich et al., 2019). Some of these biomarkers correspond to possible target immune responses for future vaccines, but their causal effects on HIV-1 infection have not been described. The methods derived in this paper are motivated by this problem, where the biomarkers are measured with error and their effects are likely subject to confounding, but there is no supplemental data available.

This paper proceeds as follows. In Section 2 notation and the estimand are defined. In Section 3 assumptions are stated and three estimators that adjust for concurrent confounding and exposure measurement error using a conditional score approach are proposed. In Section 4 the proposed estimators are evaluated in a simulation study, and in Section 5 one of the estimators is applied to study two biomarkers collected in the HVTN 505 vaccine trial. Section 6 concludes with a discussion of the advantages and limitations of the proposed methods.

2. Notation and Estimand

Suppose there are m exposures/treatments of interest which may or may not be measured correctly. Let $\boldsymbol{A} = (A_1, A_2, ..., A_m)$ be a row vector denoting the true exposures and $\boldsymbol{A}^* = (A_1^*, A_2^*, ..., A_m^*)$ be the corresponding set of measured exposures. Suppose only the first j exposures are measured with error, such that $A_k = A_k^*$ for $k > j$. For example, in the HVTN 505 trial, a biomarker of interest A is antibody-dependent cellular phagocytosis

activity. This biomarker was not observed exactly, but an imperfect phagocytic score A^* was measured using flow cytometry analysis. Exposures subject to measurement error are assumed to be continuous random variables, while exposures known to be correctly measured may be continuous or discrete. Let Y be the outcome of interest. Define $Y(\boldsymbol{a})$ to be the potential outcome under exposure $\boldsymbol{A} = \boldsymbol{a} = (a_1, a_2, ..., a_m)$. Assuming $j \geqslant 1$, there is at least one continuous exposure and each individual has infinite potential outcomes. Let $\boldsymbol{L} = (L_1, L_2, ..., L_p)$ represent a vector of baseline covariates measured prior to exposure. Assume that n i.i.d. copies of the random variables $(\boldsymbol{L}, \boldsymbol{A}^*, Y)$ are observed.

The estimand of interest is the mean dose-response surface, namely $\mathrm{E}\{Y(\boldsymbol{a})\}$ for $\boldsymbol{a} \in \boldsymbol{\mathcal{A}}$ where $\boldsymbol{\mathcal{A}}$ represents the m-dimensional space of exposure values of interest. For example, with one exposure, $\mathrm{E}\{Y(\boldsymbol{a})\}$ may be a dose response curve across a closed interval of exposure values. Each of the proposed estimators described in this paper will make assumptions that explicitly or implicitly impose restrictions on the surface. For example, the proposed IPW estimator will target the parameters of a marginal structural model (MSM) given by

$$g(\mathrm{E}[Y(\boldsymbol{a})]) = \gamma(1, \boldsymbol{a})^T \tag{1}$$

where g is a link function for an exponential family density, e.g., $g(\cdot) = \mathrm{logit}(\cdot)$. The MSM parameter $\gamma = (\gamma_0, ..., \gamma_m)$ is a row vector of length $m + 1$ which quantifies the effects of the exposures on the outcome. Thus the IPW estimator assumes the dose-response surface is linear on the scale of the link function. The linearity assumption facilitates use of the conditional score method (CSM) described in the next section; estimation procedures for non-linear MSM specifications are considered in Web Appendix D. The g-formula and doubly-robust estimators will not make this linearity assumption, but they will make an outcome regression assumption that implicitly places restrictions on the dose-response surface.

In general, let $F(x)$ denote $P(X \leqslant x)$ and $F(w|X)$ denote $P(W \leqslant w|X)$ for any random variables X and W. The proposed methods in Section 3 rely on the CSM modeling

assumptions described in Section 3.1 as well as a standard set of assumptions used in causal inference: (i) causal consistency, $Y = Y(\boldsymbol{a})$ when $\boldsymbol{A} = \boldsymbol{a}$; (ii) conditional exchangeability, $Y(\boldsymbol{a}) \perp\!\!\!\perp \boldsymbol{A}|\boldsymbol{L}$; and (iii) positivity, $dF(\boldsymbol{a}|\boldsymbol{l}) > 0$ for all $\boldsymbol{l}$ such that $dF(\boldsymbol{l}) > 0$. In addition, assume that the outcome and covariates are not measured with error and that there is no model mis-specification unless otherwise stated.

[Figure 1 about here.]

The methods proposed in this paper are applicable in settings such as the example directed acyclic graph (DAG) in Figure 1. The methods can accommodate the four types of exposures in the DAG: unconfounded and correctly measured, unconfounded and mismeasured, confounded and correctly measured, and confounded and mismeasured. Here and throughout, exposure measurement error is assumed non-differential with respect to the outcome, e.g., in Figure 1, $U_1, U_4 \perp\!\!\!\perp Y$ because the corresponding measured exposures A_1^* and A_4^* are colliders on all paths between the errors and the outcome. Many existing methods that adjust for both confounding and measurement biases focus on only mismeasured exposures or on a set of correctly measured exposures with one additional mismeasured exposure; the methods developed in this paper provide a more broadly usable modeling framework where all four types of exposures represented in Figure 1 can be analyzed simultaneously.

3. Methods

The proposed methods combine existing methods for (i) correcting exposure measurement error using CSM and (ii) adjusting for confounding using g-formula, inverse probability weighting, and doubly-robust techniques. To begin, CSM is briefly reviewed.

3.1 *Review of Conditional Score Method*

Suppose the conditional density of the outcome given exposures and covariates is in the exponential family (McCullagh and Nelder, 1989), i.e., $f(y|\boldsymbol{a}, \boldsymbol{l}, \Theta) = \exp[\{y\eta - \mathcal{D}(\eta)\}/\phi +$

$c(y, \phi)]$ where $\eta = \beta_0 + \boldsymbol{a}\beta_a + \boldsymbol{l}\beta_l$, $\mathcal{D}$ and c are functions, and $\Theta = (\beta_0, \beta_a^T, \beta_l^T, \phi)$ represents the parameters to be estimated. Assume a classical additive measurement error model $\boldsymbol{A}^* = \boldsymbol{A} + \epsilon_{me}$, where ϵ_{me} is multivariate normal with mean zero and covariance matrix Σ_{me}. To account for $A_{j+1}, ..., A_m$ being correctly measured, assume Σ_{me} has the block diagonal form

$$\Sigma_{me} = \begin{bmatrix} \Sigma_e & 0_{j,m-j} \\ 0_{m-j,j} & 0_{m-j,m-j} \end{bmatrix}$$

where Σ_e is the measurement error covariance matrix for exposures $A_1, ..., A_j$ and $0_{a,b}$ denotes an $a \times b$ matrix of zeros.

CSM leverages the fact that $\boldsymbol{\Delta} = \boldsymbol{A}^* + Y(\Sigma_{me}\beta_a)^T/\phi$ is a sufficient statistic for $\boldsymbol{A}$. Furthermore, the conditional density of the outcome given covariates $\boldsymbol{L} = \boldsymbol{l}$ and sufficient statistic $\boldsymbol{\Delta} = \boldsymbol{\delta}$ is in the exponential family with $\eta_* = \beta_0 + \boldsymbol{\delta}\beta_a + \boldsymbol{l}\beta_l$, $\mathcal{D}_*(\eta_*, \phi, \beta_a^*) = \phi \log\left[\int \exp\{y\eta_*/\phi + c_*(y, \phi, \beta_a^*)\}dy\right]$, and $c_*(y, \phi, \beta_a^*) = c(y, \phi) - \frac{1}{2}(y/\phi)^2\beta_a^*$, where $\beta_a^* = \beta_a^T\Sigma_{me}\beta_a$. This implies that the score equations from the likelihood conditional on $\boldsymbol{\Delta}$ yield consistent estimators for β_a that only depend on the observed data. The form of the score equations depends on the outcome model specification; the general form is

$$\psi(Y, \boldsymbol{L}, \boldsymbol{A}^*, \Theta) = \begin{bmatrix} \{Y - \mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta})\}(1, \boldsymbol{L}, \boldsymbol{\Delta})^T \\ \phi - \{Y - \mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta})\}^2/\{\mathrm{Var}(Y|\boldsymbol{L}, \boldsymbol{\Delta})/\phi\} \end{bmatrix} \tag{2}$$

where $\mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta}) = \partial\mathcal{D}_*/\partial\eta_*$ and $\mathrm{Var}(Y|\boldsymbol{L}, \boldsymbol{\Delta}) = \phi\partial^2\mathcal{D}_*/\partial\eta_*^2$. This procedure can be extended to handle interaction terms (Dagalp, 2001) by letting $\eta = \beta_0 + \boldsymbol{a}\beta_a + \boldsymbol{l}\beta_l + \boldsymbol{a}\beta_{al}\boldsymbol{l}^T$, where β_{al} is an $m \times p$ matrix of interaction parameters, $\boldsymbol{\Delta} = \boldsymbol{A}^* + Y\{\Sigma_{me}(\beta_a + \beta_{al}\boldsymbol{L}^T)\}^T/\phi$, and the appropriate elements of β_{al} are constrained to zero such that only relevant interactions are included in the model. Then one can derive score equations of the same form as (2), but replacing $(1, \boldsymbol{L}, \boldsymbol{\Delta})^T$ with $(1, \boldsymbol{L}, \boldsymbol{\Delta}, \boldsymbol{L} \otimes \boldsymbol{\Delta})^T$ where $\boldsymbol{L} \otimes \boldsymbol{\Delta}$ is a row vector containing the product of the k^{th} element of $\boldsymbol{L}$ and the r^{th} element of $\boldsymbol{\Delta}$ if and only if the (r, k) element of β_{al} is not constrained to be zero.

3.2 *G-formula CSM Estimator*

The first proposed method combines the g-formula with CSM. When there is no measurement error, the g-formula estimator is $\hat{\mathrm{E}}\{Y(\boldsymbol{a})\} = n^{-1}\sum_{i=1}^{n}\hat{\mathrm{E}}(Y_i|\boldsymbol{A}=\boldsymbol{a},\boldsymbol{L}_i)$ where the predicted mean outcomes $\hat{\mathrm{E}}(Y_i|\boldsymbol{A}=\boldsymbol{a},\boldsymbol{L}_i)$ are typically estimated using parametric models. To accommodate exposure measurement error, the proposed g-formula CSM estimator takes the same form, but instead utilizes a CSM outcome model with relevant covariates and interaction terms. The g-formula CSM estimator can be computed by solving the estimating equation $\sum_{i=1}^{n}\psi_{GF-CSM}(Y_i,\boldsymbol{L}_i,\boldsymbol{A}_i^*,\Sigma_{me},\Theta_{GF})=0$, where

$$\psi_{GF-CSM}(Y,\boldsymbol{L},\boldsymbol{A}^*,\Sigma_{me},\Theta_{GF}) = \begin{bmatrix} \{Y-\mathrm{E}(Y|\boldsymbol{L},\boldsymbol{\Delta})\}(1,\boldsymbol{L},\boldsymbol{\Delta},\boldsymbol{L}\otimes\boldsymbol{\Delta})^T \\ \phi-\{Y-\mathrm{E}(Y|\boldsymbol{L},\boldsymbol{\Delta})\}^2/\{\mathrm{Var}(Y|\boldsymbol{L},\boldsymbol{\Delta})/\phi\} \\ g^{-1}(\beta_0+\boldsymbol{a}\beta_a+\boldsymbol{L}\beta_l+\boldsymbol{a}\beta_{al}\boldsymbol{L}^T)-\mathrm{E}\{Y(\boldsymbol{a})\} \end{bmatrix} \quad (3)$$

and $\Theta_{GF} = (\beta_0,\beta_a^T,\beta_l^T,vec(\beta_{al}),\phi,\mathrm{E}\{Y(\boldsymbol{a})\})$. To estimate the dose response surface in practice, one can compute $\hat{\mathrm{E}}\{Y(\boldsymbol{a})\}$ for a large grid of points $\boldsymbol{a}$ in the space of interest.

3.3 *IPW CSM Estimator*

Since the CSM estimator can be obtained by solving a set of estimating equations, it's straightforward to define an IPW extension. In particular, consider an estimator of the MSM parameters in (1) that solves $\sum_{i=1}^{n}\psi_{IPW-CSM}(Y_i,\boldsymbol{L}_i,\boldsymbol{A}_i^*,\Sigma_{me},\Theta_{IPW})=0$ where

$$\psi_{IPW-CSM}(Y,\boldsymbol{L},\boldsymbol{A}^*,\Sigma_{me},\Theta_{IPW}) = \begin{bmatrix} \psi_{PS}(\boldsymbol{L},\boldsymbol{A}^*,\zeta) \\ SW\{Y-\mathrm{E}(Y|\boldsymbol{\Delta})\}(1,\boldsymbol{\Delta})^T \\ SW\left[\phi-\{Y-\mathrm{E}(Y|\boldsymbol{\Delta})\}^2/\{\mathrm{Var}(Y|\boldsymbol{\Delta})/\phi\}\right] \end{bmatrix}, \quad (4)$$

$$SW = \frac{dF(\mathbf{A}^*)}{dF(\mathbf{A}^*|\mathbf{L};\zeta)}, \quad (5)$$

$\Theta_{IPW} = (\zeta,\gamma_0,\gamma_a^T,\phi)$, $\psi_{PS}(\boldsymbol{L},\boldsymbol{A}^*,\zeta)$ is an estimating function corresponding to one or more propensity score models with parameters ζ, and γ_0 and γ_a are the MSM parameters from (1). Heuristically, weighting works by creating a pseudo-population where confounding is no longer present. For continuous exposures, this is accomplished by weighting by the inverse of

the joint density of exposures conditional on confounders. The unconditional joint density in the numerator of SW is used to stabilize the weights, but any valid density can be used in the numerator, with different choices affecting estimator efficiency. Since weighting eliminates confounding, the estimating equations of (4) are actually simpler than (2) because covariates $\boldsymbol{L}$ do not need to be included in the outcome model. Thus, the conditional expectation and variance nested in (4) are not conditional on $\boldsymbol{L}$ as they were in (2) and (3), and the equations are indexed by the MSM parameters of interest γ_0 and γ_a rather than the β parameters from equations (2) and (3). The new forms of the conditional expectation and variance follow from setting $\eta_* = \gamma_0 + \boldsymbol{\delta}\gamma_a$ in the equations in Section 3.1.

Methods for fitting MSM with multiple treatments often use weights of the form $SW = \prod_{j=1}^{m} dF(A_j^*)/dF(A_j^*|\boldsymbol{L})$, e.g., as in Hernán et al. (2001); to factorize the denominator in this way, the m exposures $A_1, \ldots, A_m$ are assumed to be independent conditional on $\boldsymbol{L}$. This assumption will be made throughout, but it may be dubious in various applications, such as when a treatment has a direct effect on another treatment or when treatments have an unmeasured common cause. For examples of estimating the joint propensity of dependent binary exposures using mixed models, see Tchetgen and VanderWeele (2012), Perez-Heydrich et al. (2014), or Liu, Hudgens, and Becker-Dreps (2016). If a subset of exposure effects are not subject to confounding, then those exposures do not need to be included in the product defining SW.

When the weights SW are known, ψ_{PS} can be removed from $\psi_{IPW-CSM}$. In practice, the weights are usually not known and must be estimated. Models used to estimate weights will be referred to as propensity models and weight components corresponding to continuous exposures will be constructed using a ratio of normal densities estimated from linear models (Hirano and Imbens, 2004). To illustrate this, first consider the setting where the true exposures are observed. For continuous exposure A_j, a model of the form

$A_j = \zeta(1, \boldsymbol{L})^T + \epsilon_{ps}$ might be assumed, where $\epsilon_{ps} \sim \mathcal{N}(0, \sigma^2_{ps})$ and in general $\mathcal{N}(\mu, \nu)$ denotes a normal distribution with mean μ and variance ν. Then based on the fitted model, the estimated conditional density $\hat{f}(A_j|\boldsymbol{L}; \hat{\zeta})$ is used to estimate $dF(A_j|\boldsymbol{L})$. An intercept-only model is used similarly to estimate the weight numerators. Other methods and more flexible choices for weight models (Naimi et al., 2014) can also be used for continuous exposures in practice.

Now consider the exposure measurement error setting, where only $\boldsymbol{A}^*$ is observed. For the IPW CSM estimator, exposure measurement error has an opportunity to introduce additional bias via the estimation of weights. Fortunately, for weight estimation the exposures are now the propensity model(s) outcomes, and measurement error is usually much less problematic for a model outcome than for model covariates (Carroll et al., 2006). In Web Appendix A, the proposed IPW CSM estimator is shown to be consistent even when the propensity models are fit using the mismeasured exposures $\boldsymbol{A}^*$ rather than the true exposures as described in the previous paragraph, under certain assumptions. Finally, weight components corresponding to discrete exposures (which are assumed to be always correctly measured) can be estimated using common practice approaches, such as logistic and multinomial regression.

3.4 *Doubly-Robust CSM Estimator*

Both g-formula and IPW methods rely on model specifications that may not be correct in practice. The g-formula provides consistent estimation of potential outcome means only when the outcome model conditional on exposures and confounders is correctly specified. Likewise, IPW estimators are consistent only when the propensity score models are correctly specified. In contrast, doubly-robust (DR) estimators entail specifying both propensity and outcome models, but remain consistent if one is misspecified and the other is not (Robins, Rotnitzky, and Zhao, 1994; Lunceford and Davidian, 2004; Bang and Robins, 2005).

A DR estimator for the additive measurement error setting can be derived as an extension

of the doubly-robust standardization procedure described in Vansteelandt and Keiding (2011) and Kang and Schafer (2007). Namely, one can use the standard g-formula estimator, but specify a weighted outcome regression where the weights are the inverse probability weights given in equation (5). This DR 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

$$\psi_{DR-CSM}(Y, \boldsymbol{L}, \boldsymbol{A}^*, \Sigma_{me}, \Theta_{DR}) = \begin{bmatrix} \psi_{PS}(\boldsymbol{L}, \boldsymbol{A}^*, \zeta) \\ SW\{Y - \mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta})\}(1, \boldsymbol{L}, \boldsymbol{\Delta}, \boldsymbol{L} \otimes \boldsymbol{\Delta})^T \\ SW[\phi - \{Y - \mathrm{E}(Y|\boldsymbol{L}, \boldsymbol{\Delta})\}^2 / \{\mathrm{Var}(Y|\boldsymbol{L}, \boldsymbol{\Delta})/\phi\}] \\ g^{-1}(\beta_0 + \boldsymbol{a}\beta_a + \boldsymbol{L}\beta_l + \boldsymbol{a}\beta_{al}\boldsymbol{L}^T) - \mathrm{E}\{Y(\boldsymbol{a})\} \end{bmatrix} \quad (6)$$

This estimator will be referred to as the DR CSM estimator. If the stabilized weights SW are unknown, they are estimated as described for the IPW CSM estimator; if the weights are known, ψ_{PS} can be removed from ψ_{DR-CSM}. Like the g-formula CSM estimator, the DR CSM estimator can be evaluated over a grid of values $\boldsymbol{a}$ of interest to estimate a dose-response surface.

3.5 *Handling Unknown Measurement Error Variance*

Although the proposed methods require no supplemental data and no distributional assumption on the true exposures $\boldsymbol{A}$, a priori knowledge of the measurement error covariance matrix is required. Sometimes this matrix will be known from properties of the measurement devices (e.g., some bioassays, certain widely studied survey instruments) or will be available from a prior study. Here, some guidelines are provided for analyses where this matrix is unknown.

Firstly, there are many types of studies for which the covariance matrix should be diagonal (i.e., the measurement errors should be uncorrelated). For example, biological assays run on different samples and analyzed by separate machines/researchers should have uncorrelated measurement errors, and different assays run on the same sample may often have uncorrelated or very weakly correlated measurement errors. For other types of data such as survey

responses, analysts should be more cautious, noting that response bias, recall bias, and other forms of measurement error in survey instruments may be correlated within individuals.

Some studies may have available supplemental data in the form of replicates of the potentially mismeasured variables. These replicates can be used to estimate Σ_{me} as described in Carroll et al. (2006). In particular, suppose for individual i there are k_i replicates of the mismeasured exposures, $\boldsymbol{A}^*_{i1}, ..., \boldsymbol{A}^*_{ik_i}$ with mean $\boldsymbol{A}^*_{i.}$. Then an estimator for the measurement error covariance is given by:

$$\hat{\Sigma}_{me} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{k_i}(\boldsymbol{A}^*_{ij} - \boldsymbol{A}^*_{i.})^T(\boldsymbol{A}^*_{ij} - \boldsymbol{A}^*_{i.})}{\sum_{i=1}^{n}(k_i - 1)}$$

When replicates are available, some of the existing methods described in the introduction are applicable and should give similar results to the proposed methods.

When no replicates are available and there is no prior knowledge of Σ_{me}, the proposed methods can still be used in conjunction with a sensitivity analysis. In some settings, an upper bound on the exposure measurement error variance may be assumed; for example, the variance of a correctly measured exposure (if estimated in a prior study) can act as a reasonable upper bound on measurement error variance for the corresponding mismeasured exposure. Once upper bounds are determined, inference may be repeated using the proposed methods for a range of Σ_{me} specifications to assess robustness of point estimates and confidence intervals to the degree of assumed measurement error; this procedure is more straightforward when the matrix is small and diagonal, and becomes difficult to interpret as the number of non-zero parameters grows.

3.6 *Large-Sample Properties and Variance Estimation*

In Web Appendix A, each of the three proposed estimators are shown to be consistent and asymptotically normal by showing that their corresponding estimating functions have expectation 0 (Stefanski and Boos, 2002). In addition, the DR CSM estimator is shown to be consistent when only one of the propensity and outcome models is correctly specified. Finally,

the proposed estimators are shown to be generalizations of existing popular estimators used in the no measurement error setting.

Since each proposed estimator is an M-estimator, consistent estimators of their asymptotic variances are given by the empirical sandwich variance technique. Estimating equations corresponding to the estimation of weights (and $\hat{\Sigma}_{me}$ if estimated) should be included in the estimating equation vector for each method when computing the sandwich variance estimator. Wald $100(1-\alpha)\%$ confidence intervals for the parameters of interest can then be constructed in the usual fashion. The geex R package (Saul and Hudgens, 2020) streamlines variance estimation for M-estimators and is used in the simulation and application sections.

4. Simulation Study

The performance of the proposed methods was evaluated in three simulation studies. The first simulation compared the proposed g-formula CSM approach to standard methods in a scenario where confounding and additive exposure measurement error were present. A total of 2000 data sets of $n = 800$ individuals were simulated, each with the following variables: a confounder L_1 simulated as a Bernoulli random variable with expectation 0.5, i.e., $L_1 \sim \text{Bern}(0.5)$, a second confounder $L_2 \sim \text{Bern}(0.2)$, an exposure $A \sim \mathcal{N}(2 + 0.3L_1 - 0.5L_2, 0.6)$, and an outcome $Y \sim \text{Bern}(\text{logit}^{-1}(-2 + 0.7A - 0.6L_1 + 0.4L_2 - 0.4AL_1 - 0.2AL_2))$. The exposure was subject to additive measurement error, simulated as $A^* \sim \mathcal{N}(A, 0.25)$.

Dose-response curves were estimated using four methods: (i) logistic regression specifying Y as a function of A^*, L_1, L_2 and the interactions between A^* and each covariate; (ii) CSM with the same outcome specification as the logistic regression and correctly specified measurement error variance; (iii) the g-formula with a correctly specified outcome model using the mismeasured exposures; and (iv) the proposed g-formula CSM estimator with a correctly specified outcome regression. For methods (i) and (ii), dose-response curves were estimated as $\hat{\text{E}}\{Y(a)\} = \text{logit}^{-1}(\hat{\beta}a)$ where $\hat{\beta}$ is the estimated coefficient for the

exposure from each model, in essence assuming that adjusting for L_1 and L_2 in the models will appropriately control for confounding. For the point on the true dose response curve $\mathrm{E}\{Y(a=3)\}$, the average bias, average estimated standard error, empirical standard error, and percentage of estimated confidence intervals covering the true value were estimated across the 2000 simulation runs.

Biases of the estimated dose-response curves are displayed in Figure 2. Only the proposed g-formula CSM estimator was approximately unbiased. In contrast, the other three methods performed poorly with high bias across nearly the entire range of exposure values evaluated. This is expected because these methods do not adjust for both confounding and measurement error. Likewise, only the proposed g-formula CSM estimator was unbiased with corresponding confidence intervals achieving nominal coverage for the potential outcome mean at $a = 3$ on the dose-response curve, as summarized in Table 1.

[Figure 2 about here.]

The second simulation study compared the proposed IPW CSM approach to standard methods. Let $\mathrm{Exp}(\lambda)$ refer to an exponential random variable with expectation λ. A total of 2000 data sets of $n = 800$ individuals were simulated, each with the following variables: a confounder $L \sim \mathrm{Exp}(3)$, first true exposure $A_1 \sim \mathcal{N}(4{+}0.9L, 1.1)$, second true exposure $A_2 \sim \mathcal{N}(2.5, 0.7)$, third true exposure $A_3 \sim \mathcal{N}(1.4{+}0.5, 0.6)$ and an outcome $Y \sim \mathrm{Bern}(\exp(-1.7+0.4A_1 - 0.4A_2 - 0.6A_3 + 0.7L - 0.6A_1L - 0.9A_3L - \log(\lambda/(\lambda - 0.7 + 0.6A_1 + 0.9A_3)))/(1 + \exp(-1.7{+}0.4A_1{-}0.4A_2{-}0.6A_3)))$ such that $\mathrm{logit}[\mathrm{E}\{Y(a_1, a_2, a_3)\}] = \gamma_0{+}\gamma_1a_1{+}\gamma_2a_2{+}\gamma_3a_3 = -1.7 + 0.4a_1 - 0.4a_2 - 0.6a_3$. The third exposure A_3 was allowed to be correctly measured, while A_1 and A_2 were subject to additive measurement error simulated as $A_1^* \sim \mathcal{N}(A_1, 0.36)$ and $A_2^* \sim \mathcal{N}(A_2, 0.25)$ (i.e., the measurement error covariance matrix was diagonal).

[Table 1 about here.]

The MSM parameters of interest γ_1, γ_2, and γ_3 were estimated from the observed data

$\{Y, A_1^*, A_2^*, A_3, L\}$ using four methods: (i) logistic regression specifying Y as a function of A_1^*, A_2^*, A_3, L and the interactions between A_1^* and L as well as A_3 and L; (ii) the CSM method with the same outcome specification as the logistic regression and correctly specified measurement error variances; (iii) an IPW estimator where a correctly specified propensity model was fit for both A_1^* and A_3 and the product of the resulting inverse probability weights were used to weight a logistic regression of Y on the three observed exposures; and (iv) the proposed IPW CSM estimator using correctly specified measurement error variances and weights from correctly specified propensity models. Methods (i) and (ii) estimate γ_1, γ_2, and γ_3 by the main effect coefficient estimates of the respective exposures, assuming that adjusting for L in the models will appropriately control for confounding.

[Table 2 about here.]

The average bias, average estimated standard error, empirical standard error, and percentage of confidence intervals covering the true values across the 2000 simulation runs are presented in Table 2. For γ_2, the parameter corresponding to the mismeasured but unconfounded exposure, methods (ii) and (iv) achieved near-zero average bias and approximately nominal coverage, while methods (i) and (iii) were biased and had lower coverage. For γ_3, the parameter corresponding to the confounded but correctly measured exposure, methods (iii) and (iv) achieved low bias and about nominal coverage, while methods (i) and (ii) performed poorly. For γ_1, the parameter corresponding to the mismeasured and confounded exposure, only the method proposed in this paper (iv) had approximately 0 bias and nominal coverage.

The third simulation study compared the proposed g-formula and IPW CSM approaches to the proposed DR CSM estimator under various model specifications. In particular, 2000 datasets of $n = 2000$ individuals were simulated with the following variables: a confounder $L_1 \sim \text{Binom}(0.5)$, a second confounder $L_2 \sim \mathcal{N}(1, 0.5)$, an exposure $A \sim \mathcal{N}(2 + 0.9L_1 - 0.6L_2, 1.1)$, and a continuous outcome $Y \sim \mathcal{N}(1.5 + 0.7A + 0.9L_1 - 0.7AL_1 - 0.6L_2 + 0.4AL_2)$

such that the assumptions of all three methods hold given correct model specifications. The methods were compared in performance estimating the coefficient for a in the marginal structural model resulting from these distributional assumptions. The exposure A was subject to additive measurement error simulated as $A^* \sim \mathcal{N}(A, 0.16)$. Then the three approaches were compared under scenarios where only the propensity model was correctly specified, only the outcome regression was correctly specified, and where both were correctly specified. The propensity model was mis-specified by not including the confounder L_1 and the outcome regression was mis-specified by leaving out L_1 and the interaction between A and L_1.

[Table 3 about here.]

The simulation results are presented in Table 3 and match the theoretical results described in Section 3.4. Namely, when only the propensity score model was specified correctly, the IPW estimator performed well, but the g-formula estimator was subject to substantial bias and undercoverage. Likewise when only the outcome model was specified correctly, the g-formula estimator performed well, but the IPW estimator was biased and had lower than nominal coverage. However, the doubly-robust estimator achieved low bias and approximately nominal coverage when only one of the two models was misspecified, exhibiting its namesake double-robustness property. While all simulations in this section generated data under positivity and additive measurement error, simulations of the proposed methods under violations of the positivity and additive measurement error assumptions are presented in Web Table 2 and Web Table 3.

5. Application

To illustrate the proposed methods, the DR CSM estimator was applied to data from the HVTN 505 vaccine trial. As discussed in the Introduction, the candidate HIV vaccine studied in this trial was not effective, but follow-up research described several interesting

biomarker correlates of risk. Recently Neidich et al. (2019) investigated potential mechanisms of antibody mediated prevention, finding that antibody-dependent cellular phagocytosis and antigen-specific recruitment of Fcγ receptors of several HIV-1 specific Env proteins were associated with reduced HIV-1 risk.

In this section, the primary analysis of Neidich et al. (2019) is reassessed by (i) adjusting for measured potential confounders beyond simple inclusion in an outcome regression, (ii) allowing for additive measurement error, and (iii) estimating full dose-response curves rather than coefficients from logit-linear models. Attention is focused on the log transforms of markers measuring antibody-dependent cellular phagocytosis and recruitment of FcγRIIa of the H131-Con S gp140 protein, which will be referred to as ADCP and RII. The primary analysis of Neidich et al. (2019) focused on the effect of each of these exposures individually on HIV-1 acquisition among vaccinees. For each exposure, they fit a logistic regression model and reported odds ratios for the main effect of exposure adjusting for age, BMI, race, and behavior risk, as well as CD4 and CD8 polyfunctionality scores (CD4-P and CD8-P).

In this section, the data is analyzed making the same binomial family distribution assumption for the outcome and using the proposed doubly-robust CSM estimator. To account for the two-phase sampling design used in the HVTN 505 trial, the weights in the doubly-robust estimator are multiplied by inverse probability of sampling weights, following the procedure described in Wang et al. (2009). This version of the proposed estimator is described in more detail and evaluated in a simulation study in Web Appendix B. ADCP and RII were modeled separately to match the univariate-style analysis performed in Neidich et al. (2019). For the propensity model specification, main effects for baseline covariates age, race, BMI, behavior risk, CD4-P, and CD8-P were included. For the outcome model specification, main effects for the exposure of interest, age, race, BMI, behavior risk, CD4-P, and CD8-P were included, along with interactions of the exposure with age and the exposure

with behavior risk. Based on the theoretical and empirical results in Sections 3 and 4, the estimator should be consistent if either specification is correct. Finally, each exposure was assumed to follow a classical additive measurement error model where a sensitivity analysis was performed by varying the measurement error variances from 0 to 1/6, 1/4, and 1/3 of the variance for each exposure variable when restricted to vaccinees with an immune response. Since ADCP and RII are log-transformations of strictly positive random variables, this setup is equivalent to assuming that their corresponding non-log transformed variables follow multiplicative measurement error models. Dose response curves were estimated for each exposure and each measurement error level across a range of 0.5 to 3 for ADCP and 7 to 11 for RII.

[Figure 3 about here.]

The analysis results are plotted in Figure 3. For each exposure, lower values corresponded to higher HIV risk among the trial population, in line with prior results and biological theory. Moving across panels from top to bottom in Figure 3, the assumed measurement error variance increases and the confidence regions become wider, as expected. While higher levels of each exposure appear to be protective, there is substantial uncertainty when allowing for higher magnitudes of measurement error and because of the low sample size. Thus, studies measuring these biomarkers in more participants are needed to draw stronger conclusions.

6. Discussion

In this paper, estimators are proposed which adjust for both confounding and measurement error and which are consistent without any supplemental data and without specifying strong distributional or extrapolation assumptions. The proposed methods are semi-parametric in that while parametric assumptions for the errors were made, no assumptions were made about the distributions of true exposures. Accompanying this paper, the R package causalCSME

has been developed (see Supporting Information) which includes code examples for the conditional score method, which have not previously been provided in even the standard (not causal inference focused) setting.

While the proposed methods were shown to have favorable theoretical and empirical properties, they are not without limitation. In particular, the methods require that the measurement error covariance is known or has been previously estimated. As demonstrated in Section 5, if the covariance is unknown then sensitivity analysis can be straightforward and informative if the covariance matrix is small or restricted such that it has few parameters. In addition, modeling assumptions on $\boldsymbol{A}^*$ and Y that may be implausible in some settings were necessary. However, both additive measurement error models and the generalized linear model framework are commonly used and realistic in many settings. The DR CSM estimator provides some protection from mis-specification of these models, requiring only one but not necessarily both to be correctly specified.

There are several possible extensions of the methods described in this paper. Methods that accommodate different measurement error models and more flexible outcome model specifications would be useful. Generalizations of the conditional score approach such as in Tsiatis and Ma (2004) could be adapted to the causal setting to weaken some of the parametric assumptions made in this paper. A version of this idea is proposed by Liu et al. (2017), which considers adding confounders as additional covariates in the outcome regressions of Tsiatis and Ma (2004). In the setting with no measurement error, adding confounders to outcome regressions has been shown to yield consistent estimates of causal parameters under certain assumptions, but those assumptions are more restrictive than that needed for g-methods such as IPW and the g-formula to yield consistent estimates. Assuming similar results hold in the measurement error setting, expanding the Liu et al.

(2017) approach using IPW, g-formula, and DR techniques could be a useful extension of this work.

Another possible future direction of research would be to combine g-estimation and CSM to estimate parameters of a structural nested model where at least one exposure is measured with error. In addition, while this paper expands on the conditional score estimation procedure described in Stefanski and Carroll (1987), the corrected score estimation procedure (Nakamura, 1990) also belongs to the score equation family of functional methods, can also correct for measurement error without supplemental data, and has advantages and disadvantages compared to the conditional score approach in various settings. A causal extension of the corrected-score approach would be valuable for many applications. Finally, this paper focuses on point-exposures, but conditional-score based approaches addressing measurement error for longitudinal and survival data have been described and could be extended to define new causal inference methods in those settings.

Acknowledgements

The authors thank Kayla Kilpatrick, Shaina Mitchell, Sam Rosin, Bonnie Shook-Sa, and Jaffer Zaidi for helpful comments and suggestions, as well as the investigators, staff, and participants of the HVTN 505 trial. This work was supported by NIH grant R37 AI054165.

References

Bang, H. and Robins, J. M. (2005). Doubly robust estimation in missing data and causal inference models. *Biometrics* **61,** 962–973.

Braun, D., Gorfine, M., Parmigiani, G., Arvold, N. D., Dominici, F., and Zigler, C. (2017). Propensity scores with misclassified treatment assignment: a likelihood-based adjustment. *Biostatistics* **18,** 695–710.

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

Cole, S. R., Jacobson, L. P., Tien, P. C., Kingsley, L., Chmiel, J. S., and Anastos, K. (2010). Using marginal structural measurement-error models to estimate the long-term effect of antiretroviral therapy on incident AIDS or death. *American Journal of Epidemiology* **171,** 113–122.

Dagalp, R. E. (2001). Estimators for generalized linear measurement error models with interaction terms. *Doctoral Dissertation, North Carolina State University* .

Decamp, A. C., Rolland, M., Edlefsen, P. T., Sanders-Buell, E., Hall, B., Magaret, C. A., Fiore-Gartland, A. J., Juraska, M., Carpp, L. N., Karuna, S. T., et al. (2017). Sieve analysis of breakthrough HIV-1 sequences in HVTN 505 identifies vaccine pressure targeting the CD4 binding site of Env-gp120. *PloS One* **12,** e0185959.

Edwards, J. K., Cole, S. R., Westreich, D., Crane, H., Eron, J. J., Mathews, W. C., Moore, R., Boswell, S. L., Lesko, C. R., Mugavero, M. J., et al. (2015). Multiple imputation to account for measurement error in marginal structural models. *Epidemiology* **26,** 645–652.

Fong, Y., Shen, X., Ashley, V. C., Deal, A., Seaton, K. E., Yu, C., Grant, S. P., Ferrari, G., deCamp, A. C., Bailer, R. T., et al. (2018). Modification of the association between T-cell immune responses and human immunodeficiency virus type 1 infection risk by vaccine-induced antibody responses in the HVTN 505 trial. *The Journal of Infectious Diseases* **217,** 1280–1288.

Hammer, S. M., Sobieszczyk, M. E., Janes, H., Karuna, S. T., Mulligan, M. J., Grove, D., Koblin, B. A., Buchbinder, S. P., Keefer, M. C., Tomaras, G. D., et al. (2013). Efficacy trial of a DNA/rAd5 HIV-1 preventive vaccine. *New England Journal of Medicine* **369,** 2083–2092.

Hernán, M. A., Brumback, B., and Robins, J. M. (2001). Marginal structural models to

estimate the joint causal effect of nonrandomized treatments. *Journal of the American Statistical Association* **96,** 440–448.

Hirano, K. and Imbens, G. W. (2004). The propensity score with continuous treatments. In Gelman, A. and li Meng, X., editors, *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives: An Essential Journey with Donald Rubin's Statistical Family*, chapter 7, pages 73–84. Wiley Blackwell.

Hong, H., Rudolph, K. E., and Stuart, E. A. (2017). Bayesian approach for addressing differential covariate measurement error in propensity score methods. *Psychometrika* **82,** 1078–1096.

Janes, H. E., Cohen, K. W., Frahm, N., De Rosa, S. C., Sanchez, B., Hural, J., Magaret, C. A., Karuna, S., Bentley, C., Gottardo, R., et al. (2017). Higher T-cell responses induced by DNA/rAd5 HIV-1 preventive vaccine are associated with lower HIV-1 infection risk in an efficacy trial. *The Journal of Infectious Diseases* **215,** 1376–1385.

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.

Kendall, B. E. (2015). A statistical symphony: instrumental variables reveal causality and control measurement error. In Fox, G., Negrete-Yankelevich, S., and Sosa, V., editors, *Ecological Statistics: Contemporary Theory and Application*, chapter 7. Oxford University Press.

Kuroki, M. and Pearl, J. (2014). Measurement bias and effect restoration in causal inference. *Biometrika* **101,** 423–437.

Kyle, R. P., Moodie, E. E., Klein, M. B., and Abrahamowicz, M. (2016). Correcting for measurement error in time-varying covariates in marginal structural models. *American Journal of Epidemiology* **184,** 249–258.

Liu, J., Ma, Y., Zhu, L., and Carroll, R. J. (2017). Estimation and inference of error-prone covariate effect in the presence of confounding variables. *Electronic Journal of Statistics* **11,** 480–501.

Liu, L., Hudgens, M. G., and Becker-Dreps, S. (2016). On inverse probability-weighted estimators in the presence of interference. *Biometrika* **103,** 829–842.

Lockwood, J. and McCaffrey, D. F. (2015). Simulation-extrapolation for estimating means and causal effects with mismeasured covariates. *Observational Studies* **1,** 241–290.

Lunceford, J. K. and Davidian, M. (2004). Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine* **23,** 2937–2960.

McCaffrey, D. F., Lockwood, J., and Setodji, C. M. (2013). Inverse probability weighting with error-prone covariates. *Biometrika* **100,** 671–680.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Routledge.

Naimi, A. I., Moodie, E. E., Auger, N., and Kaufman, J. S. (2014). Constructing inverse probability weights for continuous exposures: a comparison of methods. *Epidemiology* **25,** 292–299.

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

Neidich, S. D., Fong, Y., Li, S. S., Geraghty, D. E., Williamson, B. D., Young, W. C., Goodman, D., Seaton, K. E., Shen, X., Sawant, S., et al. (2019). Antibody Fc effector functions and IgG3 associate with decreased HIV-1 risk. *Journal of Clinical Investigation* **129,** 4838–4849.

Perez-Heydrich, C., Hudgens, M. G., Halloran, M. E., Clemens, J. D., Ali, M., and Emch, M. E. (2014). Assessing effects of cholera vaccination in the presence of interference. *Biometrics* **70,** 731–741.

Robins, J. M., Rotnitzky, A., and Zhao, L. P. (1994). Estimation of regression coefficients when some regressors are not always observed. *Journal of the American statistical Association* **89,** 846–866.

Saul, B. and Hudgens, M. (2020). The calculus of M-Estimation in R with geex. *Journal of Statistical Software* **92,** 1–15.

Shu, D. and Yi, G. Y. (2019). Inverse-probability-of-treatment weighted estimation of causal parameters in the presence of error-contaminated and time-dependent confounders. *Biometrical Journal* **61,** 1507–1525.

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.

Tchetgen, E. J. T. and VanderWeele, T. J. (2012). On causal inference in the presence of interference. *Statistical Methods in Medical Research* **21,** 55–75.

Tsiatis, A. A. and Ma, Y. (2004). Locally efficient semiparametric estimators for functional measurement error models. *Biometrika* **91,** 835–848.

Vansteelandt, S., Babanezhad, M., and Goetghebeur, E. (2009). Correcting instrumental variables estimators for systematic measurement error. *Statistica Sinica* **19,** 1223–1246.

Vansteelandt, S. and Keiding, N. (2011). Invited commentary: G-computation - lost in translation? *American Journal of Epidemiology* **173,** 739–742.

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.

Wu, X., Braun, D., Kioumourtzoglou, M.-A., Choirat, C., Di, Q., and Dominici, F. (2019). Causal inference in the context of an error prone exposure: air pollution and mortality.

The Annals of Applied Statistics **13,** 520–547.

Supporting Information

Web Appendices, Tables, and Figures referenced in Sections 3 and 6 are available with this paper at the Biometrics website on Wiley Online Library. All R code used in the simulations and application is available at https://github.com/bblette1/causalCSME. The data that support the findings of this study are available on Atlas at https://atlas.scharp.org.

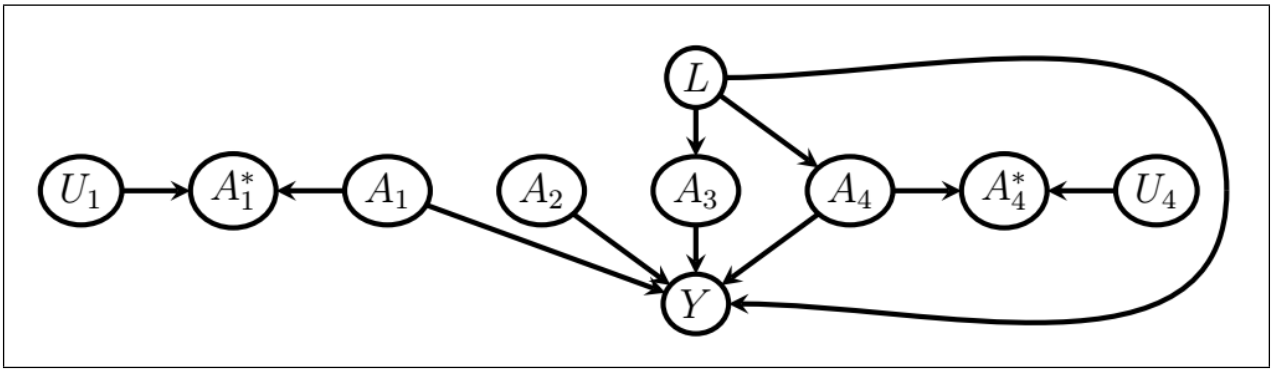

Figure 1. An example directed acyclic graph (DAG), with variables defined in Section 2 and U_1 and U_4 corresponding to measurement error. This DAG represents a scenario with $m = 4$ exposures, one each of the following: mismeasured and unconfounded (A_1), correctly measured and unconfounded (A_2), correctly measured and confounded (A_3), and mismeasured and confounded (A_4). In this case, $A_2 = A_2^*$ and $A_3 = A_3^*$ since they are both measured without error.

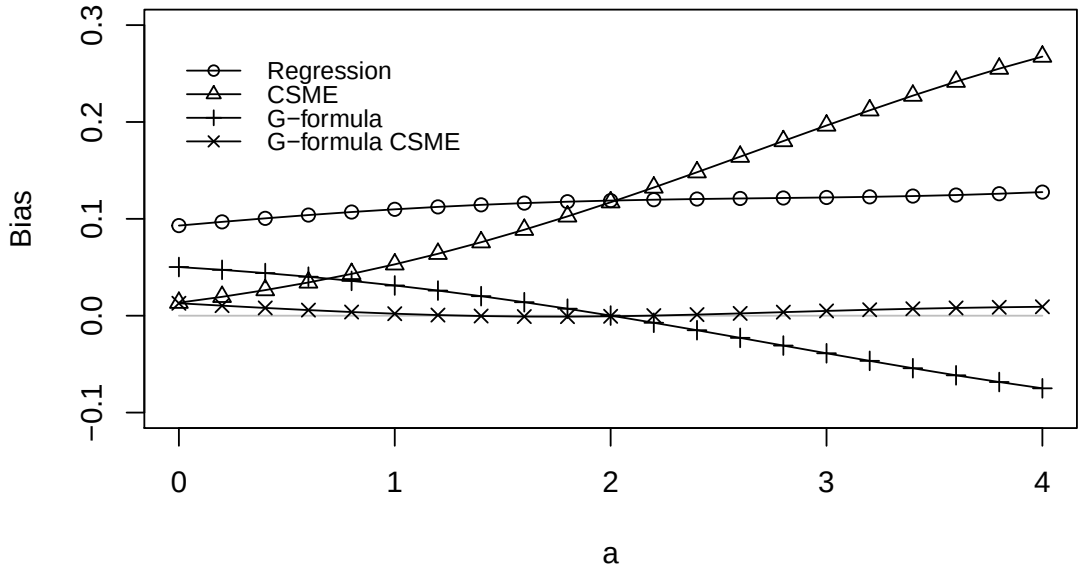

Figure 2. Estimated dose-response curve bias for each of the four methods compared in the first simulation study. Bias refers to the average bias across 2,000 simulated data sets for each method evaluated at each point on the horizontal axis $a = (0, 0.2, 0.4, ..., 4)$. The gray horizontal line corresponds to zero bias.

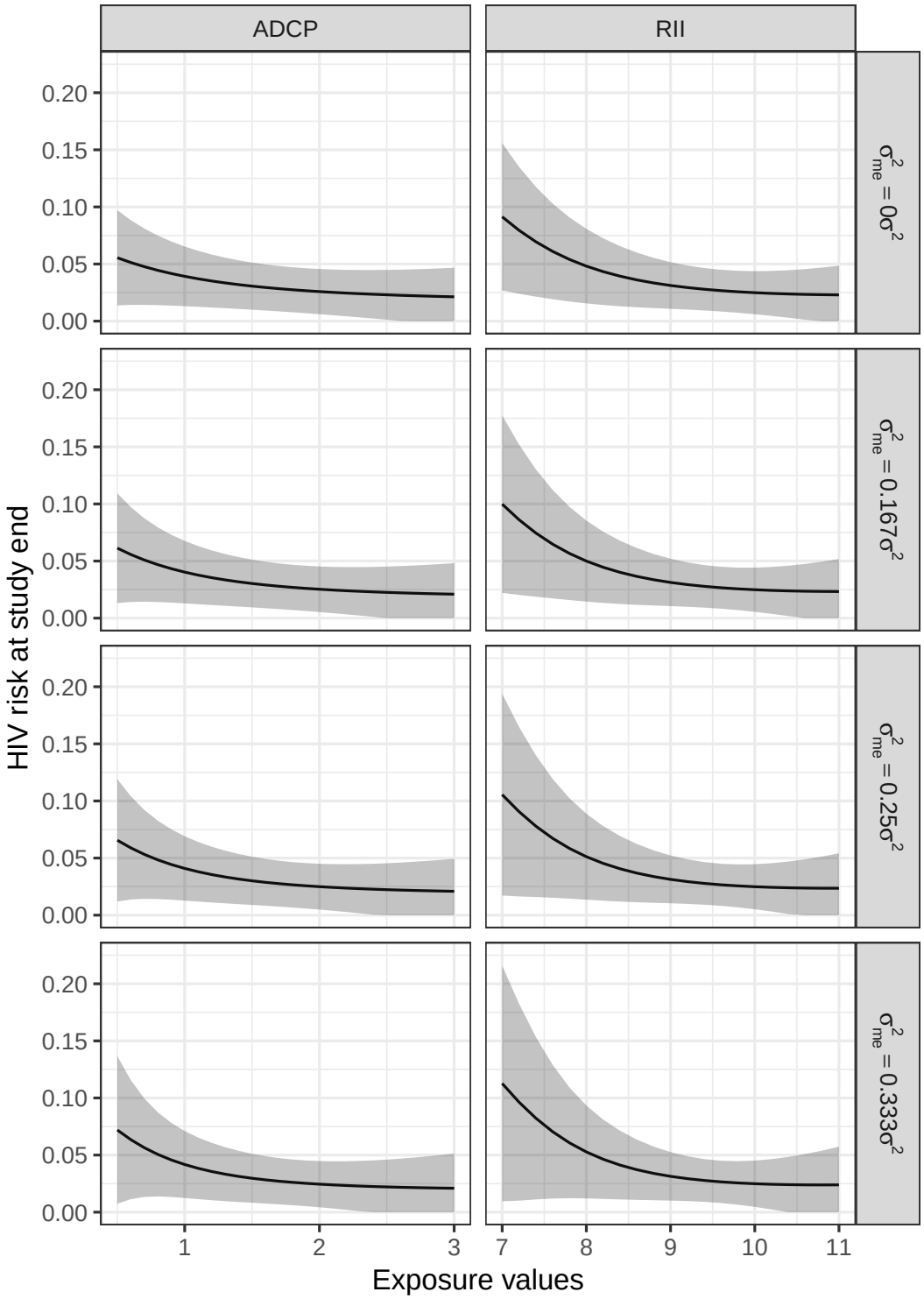

Figure 3. HVTN 505 results. Each panel shows the dose-response curve for ADCP or RII estimated by the DR CSM estimator, as well as their respective shaded confidence regions. From top to bottom, each panel reflects increasing user-specified variances of measurement error σ^2_{me} corresponding to proportions of 0, 1/6, 1/4, and 1/3 of σ^2, the estimated total exposure variances among vaccinees with immune responses. Lower confidence limits that were estimated to be negative were truncated at 0.

Table 1

Results from the first simulation study. 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.

Estimator	Bias	ASE	ESE	Cov
Regression	-51.5	18.3	18.3	19%
CSM	-21.8	28.4	28.6	87%
G-formula	-3.9	2.6	2.6	67%
G-formula CSM	0.5	4.0	4.1	95%

Table 2
Results from the second simulation study. Bias, ASE, ESE, and Cov defined as in Table 1.

Parameter	Estimator	Bias	ASE	ESE	Cov
γ_1	Regression	5.8	13.6	13.3	93%
γ_1	CSM	22.0	19.2	18.7	80%
γ_1	IPW	-9.7	9.1	9.0	80%
γ_1	IPW CSM	0.3	12.5	12.3	95%
γ_2	Regression	11.7	13.3	13.0	84%
γ_2	CSM	-3.9	21.1	20.5	95%
γ_2	IPW	13.8	13.3	12.9	80%
γ_2	IPW CSM	-0.3	20.7	20.1	95%
γ_3	Regression	10.4	27.9	27.4	92%
γ_3	CSM	9.2	28.9	27.8	92%
γ_3	IPW	0.3	19.8	19.4	94%
γ_3	IPW CSM	-0.4	20.1	19.7	95%

Table 3

Results from the third simulation study. Bias, ASE, ESE, and Cov defined as in Table 1. PS indicates the propensity score model is correctly specified; OR indicates the outcome regression is correctly specified.

Estimator	Correct Specifications	Bias	ASE	ESE	Cov
G-formula CSM	PS	-6.6	1.9	1.9	8%
IPW CSM	PS	0.0	3.2	3.1	95%
DR CSM	PS	0.0	2.6	2.6	94%
G-formula CSM	OR	0.0	1.7	1.7	94%
IPW CSM	OR	-6.3	2.0	2.0	12%
DR CSM	OR	0.1	1.7	1.7	95%
G-formula CSM	PS and OR	0.0	1.7	1.7	94%
IPW CSM	PS and OR	0.0	3.2	3.1	95%
DR CSM	PS and OR	0.1	1.9	1.9	94%