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 such strong assumptions. Despite this advantage, this method has only been adapted to perform causal inference in very limited capacity, perhaps due to its increased complexity over other functional approaches and lack of available software. 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, 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 trial evaluated a candidate HIV vaccine and stopped administering immunizations in 2013 after reaching predetermined cutoffs for futility (Hammer et al., 2013). However, subsequent analyses of trial data described 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 treatments $\boldsymbol{A} = \boldsymbol{a} = (a_1, a_2, ..., a_m)$. 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 the exposures. 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 $E\{Y(\boldsymbol{a})\}$ 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, this may be the dose response curve across a closed interval of exposure values. Each of the proposed estimators 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 framework 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.

The proposed methods in Section 3 rely on the CSME 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, $f(\boldsymbol{a}|\boldsymbol{l}) > 0$ for all $\boldsymbol{l}$ such that $dF_{\boldsymbol{L}}(\boldsymbol{l}) > 0$, where $F_{\boldsymbol{L}}$ is the cumulative distribution function of $\boldsymbol{L}$. 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.]

An example directed acyclic graph (DAG) satisfying these assumptions is presented in Figure 1 with four exposures, one each of the following: unconfounded and correctly measured, unconfounded and mismeasured, confounded and correctly measured, and confounded and mismeasured. Here and throughout the paper, exposure measurement error is assumed non-differential with respect to the outcome, i.e., $U_1, U_4 \perp\!\!\!\perp Y$ because the corresponding measured exposures ($A_1^*$ or $A_4^*$) act as colliders on all paths between the errors and the outcome. An additional contribution of this paper is that 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 studied simultaneously.

## 3. Methods

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

### 3.1 *Review of Conditional Score Approach*

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, $\boldsymbol{a}$ and $\boldsymbol{l}$ are realizations of the random variables $\boldsymbol{A}$ and $\boldsymbol{L}$, 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 $\epsilon_{me}$ is multivariate normal with mean zero and covariance matrix $\Sigma_{me}$. To account for $A_{j+1}, ..., A_m$ being correctly measured, assume $\Sigma_{me}$ 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 $\Sigma_e$ is the measurement error covariance matrix for exposures $A_1, ..., A_j$ and $0_{a,b}$ denotes an $a \times b$ matrix of zeros.

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

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

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 $\beta_{al}$ 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 appropriate elements of $\beta_{al}$ can be constrained to zero such that only relevant interactions

are included in the model. Then one can derive estimating 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 $\boldsymbol{\Delta}$ for all $r, k$ if and only if the $(r, k)$ element of $\beta_{al}$ is not constrained to be zero.

### 3.2 *G-formula CSME Estimator*

The first proposed method combines the g-formula with the CSME method. When there is no measurement error, the g-formula estimator is written as $\hat{E}\{Y(\boldsymbol{a})\} = n^{-1} \sum_{i=1}^{n} \hat{\mathrm{E}}(Y_i | \boldsymbol{A} = \boldsymbol{a}, \boldsymbol{L}_i = l_i)$ where the predicted mean outcomes $\hat{\mathrm{E}}(Y_i | \boldsymbol{A} = \boldsymbol{a}, \boldsymbol{L}_i)$ are typically estimated using a parametric model such as GLM. To address exposure measurement error, the proposed g-formula CSME estimator takes the same form, but instead estimates the predicted mean outcomes from a CSME model with relevant covariates and interaction terms in the model specification. This estimator is a solution to the estimating equation $\sum_{i=1}^{n} \psi_{GF-CSME}(Y_i, \boldsymbol{L}_i, \boldsymbol{A}_i^*, \Sigma_{me}, \Theta_{GF}) = 0$, where $\Theta_{GF} = (\beta_0, \beta_a^T, \beta_l^T, vec(\beta_{al}), \phi, E\{Y(\boldsymbol{a})\})$ and

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

where the $\beta$ parameters are contained in the conditional expectation and variance terms as previously described. To estimate the dose response surface, one can simply estimate $\hat{E}\{Y(\boldsymbol{a})\}$ for a large grid of points $\boldsymbol{a}$ in the space of interest. In practice, the outcome regression model specification will place some restrictions on the dose-response surface.

### 3.3 *IPW CSME Estimator*

Since the CSME 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 $\sum_{i=1}^{n} \psi_{IPW-CSME}(Y_i, \boldsymbol{L}_i, \boldsymbol{A}_i^*, \Sigma_{me}, \Theta_{IPW}) = 0$ where

$\Theta_{IPW} = (\beta_0, \beta_a^T, \beta_l^T, vec(\beta_{al}), \phi, \gamma_1, ..., \gamma_m)$,

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

$$SW = \frac{f(A_1, A_2, ..., A_n)}{f(A_1, A_2, ..., A_n|\boldsymbol{L}_i)} = \frac{f(A_{i1})f(A_{i2}|A_{i1}), ..., f(A_{im}|A_{i1}, ..., A_{m-1})}{f(A_1|\boldsymbol{L})f(A_2|A_1, \boldsymbol{L}), ..., f(A_m|A_1, ..., A_{m-1}, \boldsymbol{L})} \quad (5)$$

and 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 such that $\hat{\gamma}$ is the solution to (4), along with an additional dispersion parameter estimate. 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.

Rather than the factorization used in (5), most MSM with multiple treatments use weights of the form $SW = \prod_{j=1}^{m} f(A_j)/f(A_j|\boldsymbol{L})$ as originally described by Hernán et al. (2001), which assumes that the treatments are independent conditional on $\boldsymbol{L}$. This assumption will be made throughout this paper, but it may be dubious in various applications, e.g., 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).

In practice, the weights $SW$ are usually not known and must be estimated. For the remainder of the paper, weights will be estimated using a ratio of normal densities; other methods (Hirano and Imbens, 2004) and more flexible choices for weight models (Naimi et al., 2014) can also be used in practice. An additional challenge arises when estimating weights in that the exposure measurement error has another opportunity to introduce bias. Fortunately, the exposures are now the propensity model(s) outcomes, and measurement error is usually much less serious for a model outcome than for model covariates (Carroll et al., 2006). In Web Appendix A, the proposed IPW CSME estimator is shown to be consistent even when

the propensity models are fit using the mismeasured exposures without correction, under certain assumptions.

### 3.4 *Doubly-Robust CSME 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 with exposures and confounders is correctly specified. Likewise, IPW estimators are consistent only when the propensity score model is 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 written similarly to the estimator described in Hirano and Imbens (2001). Namely, one can perform the g-formula using a weighted outcome regression where the weights are the inverse probability weights given in equation (5). This estimator is a solution to the estimating equation $\sum_{i=1}^{n} \psi_{DR-CSME}(Y_i, \boldsymbol{L}_i, \boldsymbol{A}_i^*, \Sigma_{me}, \Theta_{DR}) = 0$, where $\Theta_{DR} = \Theta_{GF}$ and

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

The $\beta$ parameters are contained in the conditional expectation and variance terms as before, and if the stabilized weights $SW$ are unknown, they can be estimated as described for the IPW CSME estimator. As for the g-formula CSME estimator, the proposed DR CSME estimator can be used across a grid of values $\boldsymbol{a}$ of interest to estimate a dose-response surface.

An alternative DR estimator, similar to those described in Robins (2000) and Neugebauer and van der Laan (2005), can also be considered and is given by the solution to the set of

estimating equations $\sum_{i=1}^{n} \psi_{DR-CSME-alt}(Y_i, \boldsymbol{L}_i, \boldsymbol{A}_i^*, \Theta) = 0$ where:

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

$h(\boldsymbol{a})$ is the numerator of the stabilized weights, $d\mu(\boldsymbol{a})$ is a Lebesgue measure, and $Q(\boldsymbol{A}, \boldsymbol{L})$ is a function of $\boldsymbol{A}$ and $\boldsymbol{L}$. This estimator is doubly-robust in the sense that it is consistent if either (i) correctly specified propensity models are used to estimate the weights or (ii) the outcome regression is a correctly specified CSME model, given by $Q(\boldsymbol{A}, \boldsymbol{L}) = \beta_0 + \beta_a \boldsymbol{A} + \beta_l \boldsymbol{L} + \beta_{al}(\boldsymbol{A} \otimes \boldsymbol{L}^T) - \psi^T(1, \boldsymbol{A}^T)^T$.

However, this alternative DR estimator is limited in that it targets the same MSM parameter estimand as the IPW estimator. This is problematic because the restrictions imposed on the dose-response surface by the assumed outcome regression will exclude a linear surface form for many link functions. Thus, the estimator will only truly be doubly-robust in settings where this phenomenon does not occur, such as when the outcome regression uses an identity link, or a log or probit link with no interactions between covariates and treatments. The DR estimator proposed based on the Hirano and Imbens (2001) approach is more flexible and is preferred in this setting; this estimator will be referred to as "the DR CSME estimator" for the remainder of the paper.

### 3.5 *Handling Unknown Measurement Error Variance*

Although the proposed methods require no supplemental data and no distributional assumption on the true exposures $\boldsymbol{A}$, they do require a priori knowledge of the measurement error covariance matrix. Sometimes, this will be known from properties of the measurement tools (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 taken on different samples and analyzed by separate machines/researchers should have uncorrelated measurement errors, and different assays taken 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.

Many studies do have a small amount of supplemental data, namely replicates of the potentially mismeasured variables. These replicates can be used to estimate $\Sigma_{me}$ as described in Carroll et al. (2006). Suppose there are $k_i$ replicates of the mismeasured exposures, $\boldsymbol{A}^*_{i1}, ..., \boldsymbol{A}^*_{ik_i}$ with mean for individual $i$ notated as $\boldsymbol{A}^*_{i.}$. Then an estimator for the measurement error covariance is given by:

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

noting that 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 $\Sigma_{me}$, the proposed methods can still be used in the context of a sensitivity analysis. Often an upper bound for measurement error variance of individual variables can be ascertained - it is atypical for measurement error variance to be higher than the variance of a variable itself, for example. Once upper bounds are identified or chosen, repeating the proposed methods under a few reasonable $\Sigma_{me}$ specifications yields a sensitivity analysis showing how robust the effect sizes and confidence intervals are to increasing additive measurement error; this procedure is more straightforward when the matrix is small and diagonal and becomes difficult to interpret as the number of 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 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. The geex R package (Saul and Hudgens, 2020) streamlines variance estimation for M-estimators and is used in the simulation and application sections. Wald $100(1-\alpha)\%$ confidence intervals for the parameters of interest for the IPW CSME estimator are then given by $\hat{\psi}_i \pm z_{1-\alpha/2} SE(\hat{\psi}_i)$ where $z_q$ is the q-quantile of the standard normal distribution and $SE(\hat{\psi}_i)$ is the square root of the $i^{th}$ element of the diagonal of the sandwich covariance matrix. Confidence intervals for potential outcome means estimated by the g-formula and DR CSME methods can be computed similarly. Estimating equations corresponding to the estimation of weights (and $\hat{\Sigma}_{me}$ if estimated) should be added to the estimating equation stacks for each method to achieve nominal coverage rates. If estimating a dose-response surface across a grid of exposure values, a confidence region can be obtained by combining the confidence intervals for each point on the grid (i.e., by constructing pointwise confidence bands).

## 4. Simulation Study

The performance of the proposed methods was evaluated in three simulation studies. The first simulation compared the proposed g-formula CSME approach to standard methods within 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 and distributions: a confounder 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 $Y \sim \text{Bern}(-2 + 0.7A - 0.6L_1 + 0.4L_2 - 0.4AL_1 - 0.2AL_2)$. The exposure was subject to additive measurement error, simulated as $A^* \sim \mathcal{N}(A, 0.5)$.

[Figure 2 about here.]

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) the CSME method 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 CSME estimator with a correctly specified outcome regression. Each was compared to the true causal dose-response curve resulting from the data generating assumptions made above. The proposed method is the only approach that has approximately zero bias in estimating the dose-response curve. To evaluate coverage, consider Table 1, which summarizes each method's performance in estimating $E\{Y(a = 3)\}$. 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. In addition to near zero bias, the proposed estimator also achieved nominal coverage for the potential outcome mean corresponding to the $a = 3$ point on the dose-response curve.

[Table 1 about here.]

The second simulation study compared the proposed IPW CSME approach to standard methods. A total of 2000 data sets of $n = 800$ individuals were simulated, each with the following variables and distributions: a confounder $L \sim \text{Exp}(\lambda = 3)$, first true exposure $A_1 \sim \mathcal{N}(4 + 0.9L, 1.1)$, second true exposure $A_2 \sim \mathcal{N}(1.4 + 0.5, 0.6)$, third true exposure

$A_3 \sim \mathcal{N}(2.5, 0.7)$ and an outcome $Y \sim \text{Bern}(\exp(-1.7 + 0.4A_1 - 0.6A_2 - 0.4A_3 + 0.7L - 0.6A_1L - 0.9A_2L - \log(\lambda/(\lambda - 0.7 + 0.6A_1 + 0.9A_2)))/(1 + \exp(-1.7 + 0.4A_1 - 0.6A_2 - 0.4A_3)))$ such that $\text{logit}[E\{Y(a_1, a_2, a_3)\}] = \gamma_0 + \gamma_1 a_1 + \gamma_2 a_2 + \gamma_3 a_3 = -1.7 + 0.4a_1 - 0.6a_2 - 0.4a_3$. The second exposure $A_2$ was allowed to be correctly measured, while $A_1$ and $A_3$ were subject to additive measurement error simulated as $A_1^* \sim \mathcal{N}(A_1, 0.6)$ and $A_3^* \sim \mathcal{N}(A_3, 0.5)$ (i.e., the measurement error covariance matrix was diagonal).

Four methods were compared in performance estimating $\gamma_1$, $\gamma_2$, and $\gamma_3$ using the observed data $\{Y, A_1^*, A_2, A_3^*, L\}$. These methods were: (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 $A_2$ and $L$; (ii) the CSME 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 each confounded exposure 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 CSME estimator using correctly specified measurement error variances and weights from correctly specified propensity models.

[Table 2 about here.]

The average bias, average estimated standard error, empirical standard error, and percentage of estimated confidence intervals covering the true values were estimated across the 2000 simulation runs and are presented in Table 2. For $\gamma_3$, 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 $\gamma_2$, 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 $\gamma_1$, 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 CSME approaches to the proposed DR CSME 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 held. 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 $A^* \sim \mathcal{N}(A, 0.4)$. Then the three approaches were compared in the traditional way of studying DR estimators, namely 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 3.6. 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.

## 5. Application

To illustrate the proposed methods, the DR CSME 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. A recent paper, Neidich et al. (2019), investigated potential mechanisms of antibody mediated prevention, finding that antibody-dependent cellular phagocytosis (ADCP) and antigen-specific recruitment of Fc$\gamma$ 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 reanalyzed from a different perspective to (i) adjust for measured potential confounders beyond simple inclusion in an outcome regression, (ii) allow for additive measurement error, and (iii) estimate full dose-response curves rather than coefficients from logit-linear models. In particular, attention is focused on ADCP and recruitment of Fc$\gamma$RIIa of the H131-Con S gp140 protein, which will be referred to as RII. The primary analysis of Neidich et al. (2019) focused on the effect of the log transform 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 CSME 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, and behavior risk were included. For the outcome model specification, it's possible that BMI and behavioral risk are colliders on complex and unknown causal pathways connecting the exposures and outcome. For example, if age affects the exposures, but its effect on the outcome is fully mediated through behavioral risk, then it would be inappropriate to include both age and behavioral risk in the model. Thus, an outcome regression was specified with main effects for the exposure, age, race, CD4-P, and CD8-P and interactions between the exposure of interest with age and race. 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 the variances of measurement error were varied from 0 to 0.3 in increments of 0.1 to perform a sensitivity analysis (0.3 is about half of the variance for each exposure variable when restricted to vaccinees with an immune response and acts as a reasonable upper bound for measurement error). Since ADCP and RII are strictly positive random variables, this setup is equivalent to assuming that their corresponding non-log transformed variables follow multiplicative measurement error models.

[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, the assumed measurement error variance increases and the confidence regions become much wider, as expected. However, there appears to be strong evidence that higher levels of each exposure are protective, even when accounting for a high degree of measurement error. This provides further evidence that future vaccine candidates designed to elicit strong ADCP and RII responses may confer some protection.

The effect of low levels of RII are difficult to interpret, since only a small number of subjects exhibited this immune response.

## 6. Discussion

In this paper, estimators were 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, the true exposures were treated as nuisance parameters and no assumptions were made about their distribution. Another notable contribution of this paper is to provide an R package with 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 only work without supplemental data if the covariance of measurement error is known or has been previously estimated. As demonstrated in Section 5, if the covariance is unknown then sensitivity analysis can be straightforward and highly informative if the covariance matrix is small or restricted such that it has few parameters. In addition, in order to perform inference without making strong assumptions on $\boldsymbol{A}$, modeling assumptions on $\boldsymbol{A}^*$ and $Y$ that may be implausible in some settings were necessary. However, additive measurement error models are realistic for many variables and the GLM outcome framework is an extremely common statistical analysis procedure used in many fields.

There are several possible extensions of the methods described in this paper. Methods that accommodate different measurement error model forms and more flexible outcome model specifications would be useful. Generalizations of the conditional score approach to broader semiparametric frameworks such as that described in Tsiatis and Ma (2004) could be adapted

to the causal setting to address this and weaken some of the parametric assumptions made in this paper. A version of this idea is proposed by Liu et al. (2017), but the authors take an outcome regression approach to adjusting for confounding, which is often insufficient for causal inference.

Another logical extension would be a g-estimation based estimator that builds off the CSME method 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 described in 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-exposure data, but conditional-score based approaches addressing measurement error and data sparsity for longitudinal and survival data have been described and could be extended to define new causal inference methods in those settings.

This paper adds to a growing literature on addressing measurement error in causal inference problems. No measurement error of any variables is a key assumption in drawing causal inference, but is often left implicit in analyses despite being critical to the identification of causal estimands. Indeed, exposure measurement error is common in observational studies where investigators lack control of the exposure variables - the same conditions in which investigators worry about confounding bias. Methodological contributions at this intersection are needed to facilitate simultaneous adjustment for these prevalent issues.

## 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.

## Data Availability Statement

(Will edit this after talking to Peter) The data that support the findings of this study are openly available in [repository name] at [URL], reference number [reference number]... (old) The data for HVTN 505 is publicly available at https://atlas.scharp.org/.

## 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 (Cambridge, Mass.)* **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. (2001). Estimation of causal effects using propensity score weighting: An application to data on right heart catheterization. *Health Services and Outcomes Research Methodology* **2,** 259–278.

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.

Kendall, B. E. (2015). A statistical symphony: instrumental variables reveal causality and control measurement error. In G.A. Fox, S. N.-Y. 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.

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.

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

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, Articles* **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.

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.

## Supplementary Materials

The Web Appendix is 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.

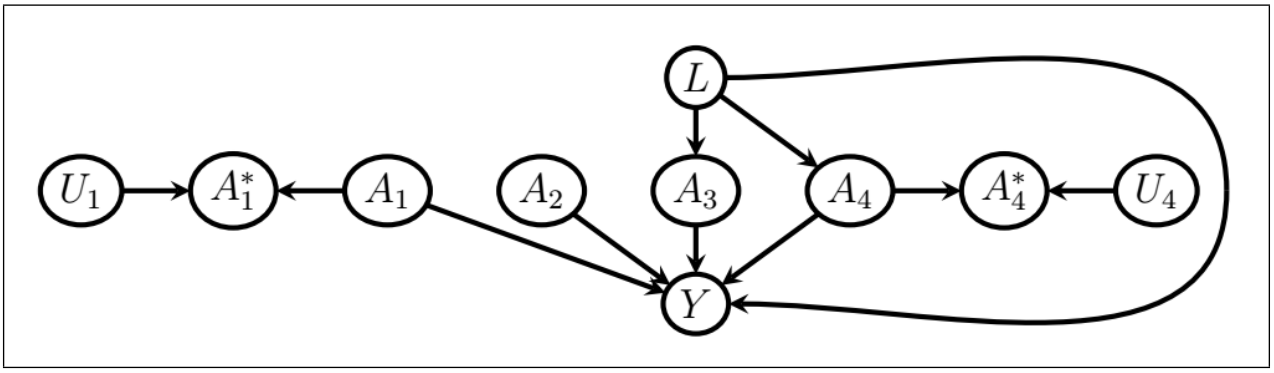


**Figure 1.** An example directed acyclic graph (DAG), with variables defined in section 2 and $U$ variables corresponding to measurement error. This DAG represents a scenario with $m = 4$ exposures, one each of the following: mismeasured and unconfounded ($A_1$), correctly measured and unconfounded ($A_2$), 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.

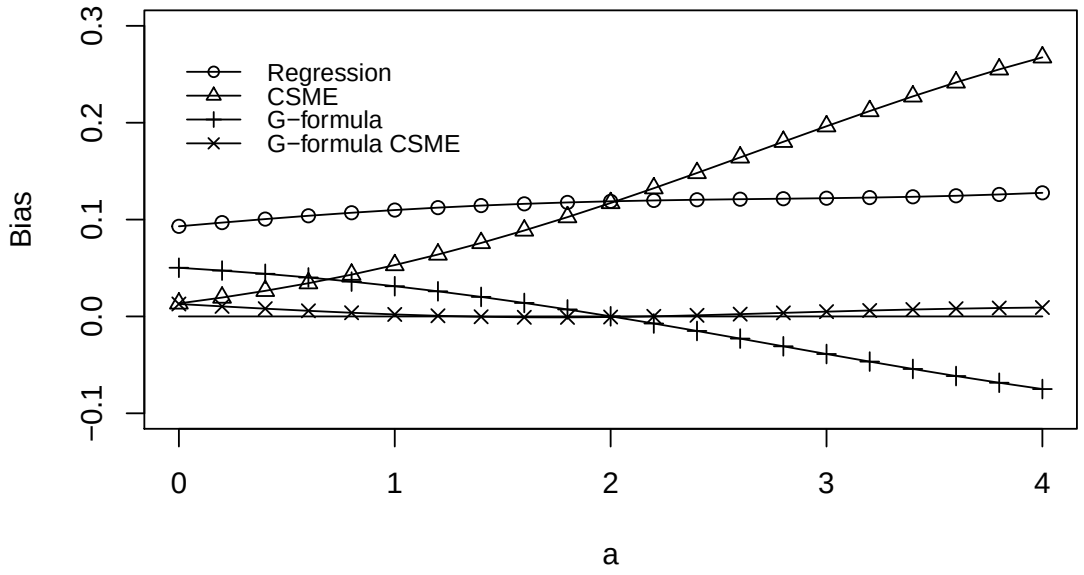


**Figure 2.** Estimated dose-response curve bias for each of the four methods compared in the first simulation study. Bias refers to the average bias across 2,000 simulated data sets for each method evaluated at each point on the horizontal axis $a = (0, 0.2, 0.4, ..., 4)$. For the sample size of 800 used in each simulation iteration, a typical support for the exposure would cover the interval from 0 to 4, with most values lying between 0.5 and 3.5.

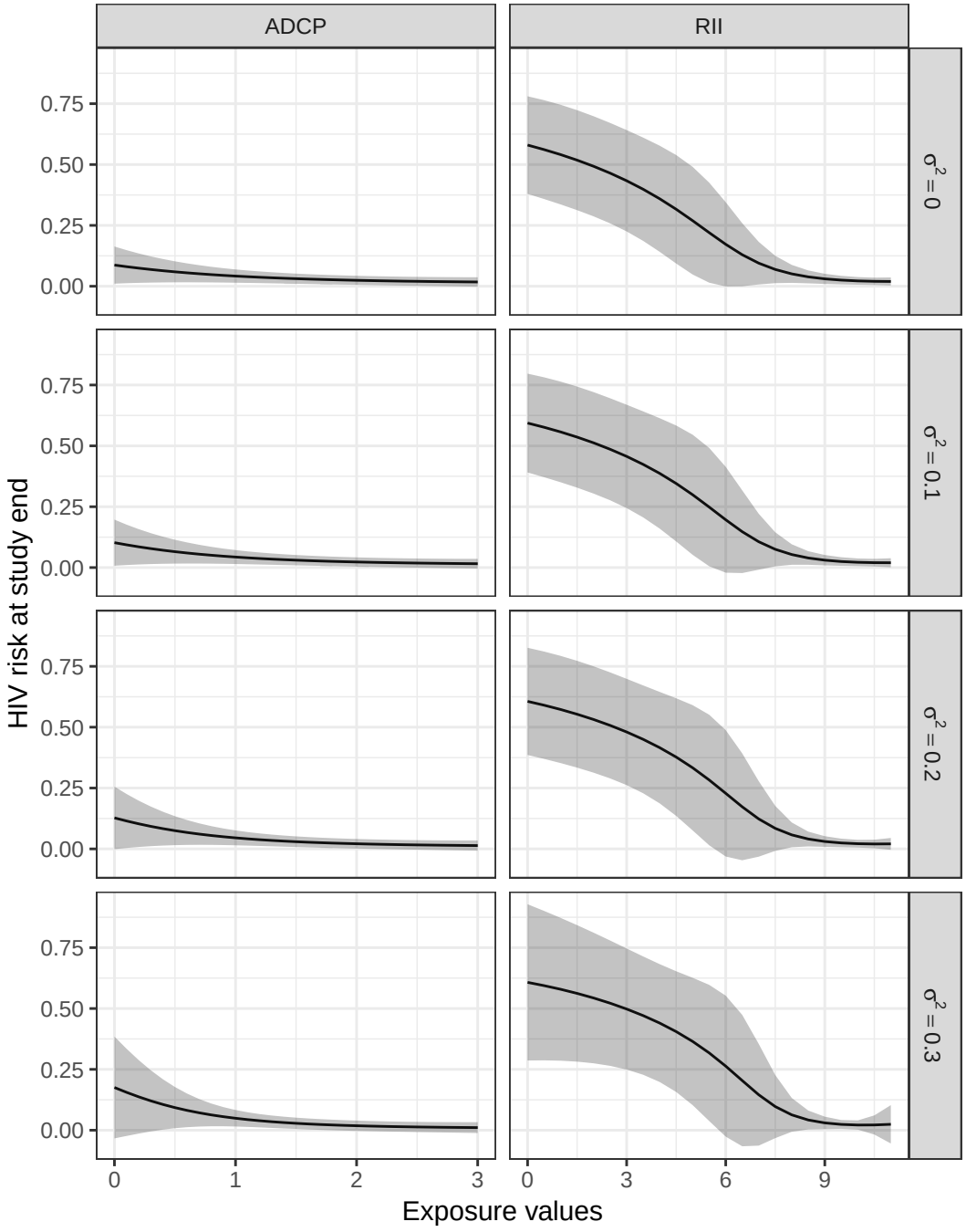


**Figure 3.** HVTN 505 results. Each panel shows the dose-response curve for ADCP or RII estimated by the DR CSME estimator, as well as their respective shaded confidence regions. From top to bottom, each panel reflects increasing user-specified variances of measurement error $\sigma^2$. The variances of 0, 0.1, 0.2, and 0.3 correspond to proportions of the observed exposure variances (among vaccinees with immune responses) of approximately 0, 1/6, 1/3, and 1/2.

**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 | Coverage |
|---|---|---|---|---|
| Regression | -51.5 | 18.3 | 18.3 | 19% |
| CSME | -21.8 | 28.4 | 28.6 | 87% |
| G-formula | -3.9 | 2.6 | 2.6 | 67% |
| G-formula CSME | 0.5 | 4.0 | 4.1 | 95% |

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

| | $\gamma_1$ | | | | $\gamma_2$ | | | | $\gamma_3$ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Estimator | Bias | ASE | ESE | Cov | Bias | ASE | ESE | Cov | Bias | ASE | ESE | Cov |
| Regression | 5.8 | 13.6 | 13.3 | 93% | 10.4 | 27.9 | 27.4 | 92% | 11.7 | 13.3 | 13.0 | 84% |
| CSME | 22.0 | 19.2 | 18.7 | 80% | 9.2 | 28.9 | 27.8 | 92% | -3.9 | 21.1 | 20.5 | 95% |
| IPW | -9.7 | 9.1 | 9.0 | 80% | 0.3 | 19.8 | 19.4 | 94% | 13.8 | 13.3 | 12.9 | 80% |
| IPW CSME | 0.3 | 12.5 | 12.3 | 95% | -0.4 | 20.1 | 19.7 | 95% | -0.3 | 20.7 | 20.1 | 95% |

**Table 3**

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

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