Published in final edited form as:

Stat Med. 2018 December 10; 37(28): 4126–4141. doi:10.1002/sim.7922.

Efficiency and Robustness of Causal Effect Estimators When Noncompliance is Measured with Error

Jeffrey A. Boatman^{1,*}, David M. Vock¹, Joseph S. Koopmeiners¹

¹Division of Biostatistics, School of Public Health, University of Minnesota, Minneapolis, Minnesota.

Abstract

Estimating causal effects from randomized controlled trials is often complicated due to participant noncompliance to randomized treatment. Although there are a variety of methods to estimate causal effects in the presence of noncompliance, they generally make the assumption that noncompliance is measured without error. This is frequently an untenable assumption, particularly when noncompliance is based on participant self-report. To overcome this issue, we treat compliance as an unobserved variable and show how to estimate the probability of compliance given a biomarker of treatment and the other observed data. We present inverse probability weighted estimators, regression-based estimators, and a doubly-robust augmented estimator that rely on the estimated probability of compliance rather than an indicator of compliance. We investigate the finite-sample properties of the estimators and their efficiency and robustness under correctly specified or misspecified models, and we apply the estimators to a recently completed trial of very low nicotine content cigarettes.

Keywords

Causal Inference; Inverse Probability Weighting; Measurement Error; Noncompliance; Randomized Controlled Trials

1. Introduction

Randomized controlled trials (RCTs) are the gold standard for establishing causation in research. By randomly assigning subjects to a treatment group, systematic differences between the groups are eliminated by design, and differences observed between treatment groups can be attributed to the intervention. Unfortunately, participant noncompliance to randomized treatment is a common problem in human subject RCTs. Noncompliance complicates assessing the effect of an intervention because, although treatment assignment is randomized, the treatment received is not. Specifically, compliance is often confounded, that is, patient characteristics are associated with compliance and with the outcome. These

*Correspondence to:

Software

The R code for the simulation and an example data set with analysis is available for download at <https://github.com/jeffrey-boatman/efficiency-robustness>.

characteristics, rather than the treatment effect, may be the cause of observed differences between treatment groups among those observed to be fully compliant.

Intention to treat (ITT) estimators, which analyze data for all participants according to randomized treatment group regardless of compliance to assigned treatment, are frequently recommended as the primary analysis because they can provide an unbiased estimator of the effect of a treatment as it will be used in practice [1]. However, this effect may be different than the effect that would be observed if all participants were to be compliant, that is, the causal effect of the treatment [2]. The causal effect is frequently of significant interest as it may elucidate biological mechanisms or provide investigators and policy experts an estimated treatment effect if compliance can be increased or enforced by regulation.

Several statistical methods have been proposed to estimate the causal effect of treatment in RCTs when subjects may not be fully compliant. Lunceford and Davidian [3] outlined three general and widely-used approaches, 1) inverse probability weighted estimators, 2) regression-based estimators, and 3) augmented inverse probability weighted estimators, to estimate causal effects when the treatment received is not randomized. Inverse probability of compliance weighted (IPCW) estimators [4] directly model the probability of compliance given the confounders. Under the assumption of no unmeasured confounders [5], and assuming that the model for compliance is correct, the weights create a pseudo-population in which the treatment is independent of measured confounder; and, therefore, IPCW estimators are consistent estimators for the causal effect. Alternatively, if a regression model for the conditional mean of the response given confounders, treatment group, and compliance status is known or can be consistently estimated, this model can be used to estimate the within-treatment group counterfactual mean assuming complete compliance by averaging the (estimated) conditional mean of the response at each level of the observed confounders for all participants assuming full compliance and assignment to a particular treatment group [3]. Augmented estimators combine IPCW and regression estimators by “augmenting” the IPCW estimator with a regression component. When the regression model is correctly specified, augmented estimators are more efficient than the IPCW estimators. Additionally, these estimators have the appealing property of being doubly-robust: if either the regression model or the model for the probability of compliance is correct, then the estimator is consistent for the causal effect even if the other model is misspecified [6].

The methods described above share the common implicit assumption that compliance is measured without error. This is often an unrealistic assumption, particularly when compliance is based on participant self-report. Specifically, we are motivated by data collected as part of the Center for the Evaluation of Nicotine in Cigarettes, Project 1 (CENIC-p1), a 6-week randomized trial of very low nicotine content (VLNC) cigarettes on cigarette consumption, dependence and other measures of smoking behavior [7]. Participants were randomized to either their usual brand cigarettes or experimental cigarettes with a range of nicotine content and were instructed to smoke only those cigarettes provided by the study (i.e., “study” cigarettes). In this study, noncompliance occurs when subjects use commercially-available tobacco products not supplied by the study. Subjects self-reported the number of study and “non-study” cigarettes smoked daily. Although 61% of study participants randomized cigarettes with the lowest nicotine content reported smoking only

study cigarettes during the final week of the trial, biomarkers of nicotine exposure (e.g., total nicotine equivalent or TNE, a biomarker of short-term nicotine exposure that measures most nicotine metabolites) were inconsistent with smoking only study cigarettes for many participants. This indicates that self-reported compliance is not a reliable indicator of actual compliance. Although we are motivated by an RCT evaluating VLNC cigarettes, investigators in other contexts often rely on imperfect measures of compliance, such as the number of pills remaining in a bottle returned to the pharmacy in a pharmaceutical trial, where other, more precise, measures of compliance are also collected. In these contexts, any consistent causal estimator must overcome the problems associated with relying on imperfect measures of compliance.

One possible solution is to take a measurement error approach and develop a causal estimator where self-reported compliance is assumed to be a surrogate for actual compliance. Several papers have developed methods to estimate causal mediation effects when the mediator may be measured with error or misclassified [8, 9]. In this context, compliance is a mediator of the effect of assigned treatment and the causal effect of treatment is equivalent to the controlled direct effect. However, these approaches typically make strong assumptions concerning the measurement error/misclassification mechanism (e.g., nondifferential measurement error or misclassification) which may not be reasonable in the context of RCTs including CENIC-p1, because subjects who report noncompliance are almost always noncompliant, but self-reported compliance is frequently erroneous.

Boatman and colleagues [10] avoided measurement error assumptions and instead treated compliance as an unobserved variable. They proposed an IPCW-like estimator that, rather than weighting participants according to the inverse of the probability of compliance given the confounders, weights participants according to the product of the probability of compliance given the observed data, including biomarkers of exposure to the treatment, and the inverse of probability of compliance given confounders. Although compliance is unobserved in their framework, they discuss how to estimate the probability of compliance given the observed data by treating the distribution of the biomarker of treatment exposure as a mixture distribution composed of compliers and noncompliers. However, IPCW estimators are known to be highly variable due to instability caused by observations with large weights [11]. Developing regression-based and augmented IPCW estimators when compliance status is measured with error represent potential approaches to reducing variability, but deriving these estimators in this context is not trivial.

Similar to Lunceford and Davidian [3], we develop a series of weighted estimators, regression-based estimators, and an augmented estimator for causal inference, but our focus is on developing these estimators in the contexts of RCTs when compliance is measured with error under minimal assumptions on the measurement error process. Throughout, we assume that we possess a biomarker of treatment exposure that can be used to identify noncompliance. Such biomarkers are often available and can aid in estimation.

The remainder of the manuscript proceeds as follows. In Section 2, we describe the potential outcomes framework, the observed data, and the assumptions necessary to identify the causal effect. In Section 3, we develop weighted estimators, regression-based estimators, and

a doubly-robust augmented estimator to estimate causal effects in RCTs when compliance is measured with error. In Section 4, we present a simulation study to assess the finite-sample performance of the estimators. We apply the estimators to data from the CENIC-p1, in Section 5, and we conclude in Section 6.

2. Preliminaries and Notation

2.1. Potential Outcomes and Target of Inference

We assume that data are collected from an RCT with treatment groups $A = 1, 2, \dots, r$. Let Z indicate a measure of compliance without error, with $Z = 0$ indicating compliance and $Z = 1$ indicating noncompliance. The precise definition of Z can be tailored to the application. For example, in a pharmaceutical trial, Z may be the number of pills not taken, and in the CENIC-p1 trial Z is the number of non-study cigarettes smoked. Define the compliance indicator $C = I(Z = 0)$, and note that we treat both C and Z as unobserved data. Let $Y^*(a, z)$ be the outcome of a randomly-selected participant if, possibly contrary to fact, $A = a$ and $Z = z$. Because for each participant we do not observe all values of $Y^*(a, z)$, $Y^*(a, z)$ is known as a potential outcome. The target of inference is $\mu(a, 0) - \mu(a', 0) = E\{Y^*(a, 0) - Y^*(a', 0)\}$, the mean difference in potential outcomes between treatment groups a and a' if all participants were to be compliant. In the context of mediation literature, $\mu(a, 0) - \mu(a', 0)$ is known as the controlled direct effect [12], that is, the treatment effect when the mediator, the measure of compliance, is set at the fixed value of 0.

2.2. Observed Data

Let Y be the observed outcome for a random participant, let X be a vector of variables associated with C and Y , and let B be a biomarker associated with exposure to treatment A . For example, in pharmaceutical trials, B might be the level of drug metabolite in the blood or urine. In CENIC-p1, B is the level of TNEs in the urine. Let D indicate self-reported compliance, i.e., $D = 1$ if a participant reports compliance, and $D = 0$ if a participant reports any noncompliance. Note that we will introduce estimators that do not rely on D nor on any surrogate of C . However, if self-reported compliance data are available, these data can be incorporated into estimators depending on the scientific question of interest. For example, it may be reasonable to assume that $C = 0$ without error if $D = 0$, that is, we may assume that participants report noncompliance without error as there is usually no social desirability associated with noncompliance. For the remainder of this paper, we will assume that any self-reported measures of compliance are part of the vector X for notational simplicity.

2.3. Identifying Assumptions

We make the following standard identifying assumptions [5] to relate the observed data to the distribution of the potential outcomes. First, we assume that we have collected sufficient covariates such that compliance is conditionally ignorable. That is, we assume that C is conditionally independent of $Y^*(a, 0)$ given X and A . This is known as the no unmeasured confounders assumption. Second, we assume that for all levels X and A , there is positive probability of compliance, that is $E(C|A, X) > 0$ for all A, X . This is known as the positivity assumption. Finally, we assume that $Y = Y^*(a, 0)$ if $A = a$ and $Z = 0$, that is, if a participant is assigned treatment group a and is compliant, then her observed outcome is the same as it

would be if she were forced to be compliant. This is known as the consistency assumption. These assumptions are required assuming either the traditional case, where C is observed, or in the case where C is unobserved. Further identifying assumptions are required when C is unobserved, and these will be stated where required.

3. Causal Estimators

For each causal estimator, we describe estimation of $\mu(a, 0)$, a mean for a single treatment group. Similar estimators can be constructed for $\mu(a', 0)$, and the estimated causal contrast is taken as the difference in estimated means. We present traditional IPCW estimators, regression-based estimators, and an augmented estimator for the case where C is observed, i.e., measured without error. For each of these cases, we develop the analogous estimators for the case where C is measured with error.

3.1. Distribution of the Biomarker of Exposure

We begin by discussing a model for the relationship between the biomarker, B , and compliance status, C . This allows us to develop causal estimators while treating C as an unobserved variable and making minimal assumption on the measurement error process. Specifically, we model (1) the distribution of the biomarker given treatment group, covariates, and the response and (2) the distribution of biomarker and response given treatment group and covariates as a two-component mixture density with a component for compliers and another for non-compliers. The parameters estimated in these models are then used in the causal estimators we discuss subsequently.

The density of $B|A, X, Y$ can be written as a two-component mixture density

$$g(b|a, x, y; \xi, \alpha) = \rho(a, x, y; \alpha) \cdot f(b|a, x, y, c = 1; \xi) + \{1 - \rho(a, x, y; \alpha)\} \cdot f(b|a, x, y, c = 0; \xi), \quad (1)$$

where f is the conditional density of B given A, X, Y , and C , indexed by parameter vector ξ , and $\rho(a, x, y; \alpha) = \Pr(C = 1|A = a, X = x, Y = y, a)$, indexed by parameter vector α . The parameters for the mixture distribution in (1) can be estimated via maximum likelihood by solving the score equations

$$\sum_{i=1}^n \frac{\partial}{\partial(\xi^T, \alpha^T)} \log g(B_i|A_i, X_i, Y_i; \xi, \alpha) = 0. \quad (2)$$

Equation (2) may be difficult to solve directly, in which case we can find the maximum likelihood estimates of $(\xi^T, \alpha^T)^T$ using the EM algorithm [13].

Similarly, the joint conditional density of $B, Y|A, X$ can be written as the mixture density

$$z(b, y|a, x; \xi, \zeta, \delta) = \omega(a, x; \delta) f(b|a, x, y, c = 1; \xi) h(y|a, x, c = 1; \zeta) + \{1 - \omega(a, x; \delta)\} \cdot f(b|a, x, y, c = 0; \xi) h(y|a, x, c = 0; \zeta), \quad (3)$$

where h is the conditional density of Y given A , X , and C , indexed by parameter vector ζ , and $\omega(a, x; \delta) = \Pr(C = 1 | A = a, X = x; \delta)$, indexed by parameter vector δ . The parameters for the mixture distribution in (3) can be estimated by solving the score equations

$$\sum_{i=1}^n \frac{\partial}{\partial(\xi^T, \zeta^T, \delta^T)^T} z(B_i, Y_i | A_i, X_i; \xi, \zeta, \delta) = 0. \quad (4)$$

As in Equation (2), Equation (4) may be difficult to solve directly, and the maximum likelihood estimates of $(\xi^T, \zeta^T, \delta^T)^T$ can be found with the EM algorithm.

In some situations there may be auxiliary compliance data with known compliers, that is, participants who were known without error to be compliant. For example, there may be pharmacokinetic/pharmacodynamic data from early-phase clinical trials that can be used to discriminate compliers and non-compliers during a larger phase III trial. Boatman et al. [10] describe in detail how data from these individuals can be used in estimating parameters of the mixture distributions. We briefly summarize how to incorporate these data. For known compliers, let $K=1$, and let $K=0$ for all other participants. Let n_k denote the number of known compliers, and let $m = n + n_k$. To include known compliers in estimation of the mixture distributions, one can solve the the modified score equations

$$\sum_{i=1}^m \frac{\partial}{\partial(\xi^T, \alpha^T)^T} \left\{ (1 - K_i) \cdot \log z(B_i | X_i, Y_i; \xi, \alpha) + K_i \cdot \log f(B_i | X_i, Y_i, c = 1; \xi) \right\} = 0$$

in place of the score equations (2), and

$$\sum_{i=1}^m \frac{\partial}{\partial(\xi^T, \zeta^T, \delta^T)^T} \left\{ (1 - K_i) \cdot \log z(B_i, Y_i | X_i; \xi, \zeta, \delta) + K_i \cdot \log f(B_i | X_i, Y_i, c = 1; \xi) \right\} = 0$$

in place of the score equations (4). That is, known compliers contribute to estimation of the mixture distributions only through the information they contain about the conditional density of B . They make no other contributions to estimation of $\mu(1, 0)$.

3.2. Inverse Probability Weighted Estimators

In the case where C is observed, a traditional IPCW estimator for $\mu(a, 0)$ is the solution to the estimating equation

$$\sum_{i=1}^n \frac{I(A_i = a)C_i}{E(C_i | X_i)} \{Y_i - \mu(a, 0)\} = 0. \quad (5)$$

For the case where C is unobserved, Boatman et al. [10] proposed estimating $\mu(a, 0)$ by solving the estimating equation

$$\sum_{i=1}^n \frac{I(A_i = a)E(C_i | A_i, B_i, X_i, Y_i)}{E(C_i | A_i, X_i)} \{Y_i - \mu(a, 0)\} = 0. \quad (6)$$

The estimating function with weights $\frac{E(C|A, B, X, Y)}{E(C|A, X)}$ is similar to the traditional IPCW estimator with weights $\frac{C}{E(C|A, X)}$, but the numerator of the weights is a conditional expectation of C given the observed data rather than an indicator function.

Boatman et al. [10] showed that in expectation the estimating function in (6) equals the estimating function in (5), so provided that the models for the numerator and denominator of the weights are correctly specified and that the assumptions in Section 2.3 are met, estimators solving Equation (6) are consistent and asymptotically normal.

Although C is unobserved, $E(C|A, B, X, Y)$ and $E(C|A, X)$ can be estimated using the observed data. Using Bayes' theorem, we can write $E(C|A = a, B = b, X = x, Y = y)$, the numerator of the weights, as

$$\frac{f(b|a, x, y, c = 1; \xi)\rho(a, x, y; \alpha)}{f(b|a, x, y, c = 1; \xi)\rho(a, x, y; \alpha) + f(b|a, x, y, c = 0; \xi)\{1 - \rho(a, x, y; \alpha)\}} \quad (7)$$

Each of the components in Equation (7) is part of the mixture distribution in (1). Thus, although C is unobserved, we can estimate $E(C|A, B, X, Y)$ by substituting estimated parameters of the mixture distribution into (7). We refer to the estimator for $\mu(a, 0)$ that uses components from the mixture density in (1) to estimate the numerator of the weights as the Compliance Unsure RE-weighted (CURE) estimator. As an alternative to the factorization of $E(C|A, B, X, Y)$ in (7), we can instead write $E(C|A = a, B = b, X = x, Y = y)$ as

$$\frac{f(b|a, x, y, c = 1; \xi)h(y|a, x, c = 1; \zeta)\omega(a, x; \delta)}{f(b|a, x, y, c = 1; \xi)h(y|a, x, c = 1; \zeta)\omega(a, x; \delta) + f(b|a, x, y, c = 0; \xi)h(y|a, x, c = 0; \zeta)\{1 - \omega(a, x; \delta)\}} \quad (8)$$

Similarly, each of the components in (8) can be estimated from the mixture density in (3). We refer to the estimator for $\mu(a, 0)$ that uses components from the mixture density in (3) as the CURE+ estimator. Similar to the CURE estimator, the CURE+ estimator is consistent assuming that the models for the density in (3) are correct and the assumptions in Section 2.3 are met.

Unlike the CURE estimator, the CURE+ estimator requires specifying the conditional distribution of Y given A, X , and C . The advantage of having to specify $h(y|a, x, c; \zeta)$ is that

this may lead to greater separation between the compliers and non-compliers, allowing for easier identification of the maximum likelihood estimates using the EM algorithm. Additional advantages of this approach will become apparent in Section 3.3.

In most applications, the denominator of the weights is unknown and must be estimated. Because the denominator is between 0 and 1, we specify a regression model $E(C|A, X) = \pi(A, X; \beta) = g^{-1} \{ \beta_0 + X^T \beta_1 + I(A=2)\beta_2 + \dots + I(A=r)\beta_r \}$, where g is a link function that maps from (0, 1) to $\mathbb{R}$, such as the logit or probit link, and $\beta = (\beta_0, \beta_1^T, \beta_2, \dots, \beta_r)^T$ is a vector of unknown regression coefficients. We can estimate β by solving the score equations

$$\sum_{i=1}^n \frac{E(C_i|A_i, B_i, X_i, Y_i) - \pi_i(A_i, X_i; \beta)}{\pi_i(A_i, X_i; \beta) \{1 - \pi_i(A_i, X_i; \beta)\}} \frac{\partial \pi_i(A_i, X_i; \beta)}{\partial \beta^T} = 0. \quad (9)$$

That is, $E(C|A, B, X, Y)$ is the “response” of the regression model for $E(C|A, X)$. Note that we can estimate β simultaneously with either $(\xi^T, \alpha^T)^T$ or $(\xi^T, \zeta^T, \delta^T)^T$ by stacking Equation (9) with Equation (2) or (4), respectively. This is equivalent to replacing $E(C|A, X, B, Y)$ in Equation (9) with the estimated expectation $E(C|A, B, X, Y; \hat{\xi}_n, \hat{\alpha}_n)$ or $E(C|A, B, X, Y; \hat{\xi}_n, \hat{\zeta}_n, \hat{\delta}_n)$, where $\hat{\xi}_n$, $\hat{\alpha}_n$, $\hat{\zeta}_n$, and $\hat{\delta}_n$ are the estimates of ξ , α , ζ , and δ .

3.3. Regression-Based Estimators

As an alternative to IPCW estimation, we can instead estimate $\mu(a, 0)$ by solving the estimating equation

$$\sum_{i=1}^n \{ \eta(A = a, X_i; \gamma) - \mu(a, 0) \} = 0, \quad (10)$$

where $\eta(A, X; \gamma) = E(Y|A, X, C=1)$, and γ is a vector of regression coefficients. This is referred to as a regression-based estimator because, prior to solving Equation (10), the conditional mean model $\eta(A, X; \gamma)$ must be specified and estimated. For example, we can assume the mean model $\eta(A, X; \gamma) = \gamma_0 + X^T \gamma_1 + I(A=2)\gamma_2 + \dots + I(A=r)\gamma_r$, but other parameterizations are possible. In the case where C is observed, one can estimate the vector of regression coefficients γ by solving the estimating equations

$$\sum_{i=1}^n C_i \{ Y_i - \eta(A_i, X_i; \gamma) \} \frac{\partial \eta(A_i, X_i; \gamma)}{\partial \gamma^T} = 0 \quad (11)$$

To estimate $\mu(a, 0)$, Equations (10) and (11) can be solved jointly, which is equivalent to replacing $\eta(A = a, X; \gamma)$ in Equation (10) with $\eta(A = a, X; \hat{\gamma}_n)$.

Assuming that the model $\eta(A, X; \gamma)$ is correct, regression-based estimators are consistent estimators for $\mu(a, 0)$ and have smaller asymptotic variance than IPCW estimators [3]. Note

that regression-based estimators have a potential advantage over IPCW estimators in that they do not require specifying a model for $E(C|A, X)$. Furthermore, unlike IPCW, they do not assign weights of 0 to some participants when estimating $\mu(a, 0)$, but this comes at the cost of needing to correctly specify $\eta(A, X; \gamma)$. See [3] for further discussion of advantages and disadvantages of weighted estimators compared to regression-based estimators.

In the case where C is unobserved, both specifications of the mixture distributions in Equations (1) and (3) offer a means of estimating γ , which can then be substituted into Equation (10). Using $E(C|A, B, X, Y)$, we can estimate the regression coefficients γ by solving the estimating equations

$$\sum_{i=1}^n E(C_i | A_i, B_i, X_i, Y_i) \{Y_i - \eta(A_i, X_i; \gamma)\} \frac{\partial \eta(A_i, X_i; \gamma)}{\partial \gamma^T} = 0. \quad (12)$$

In expectation, weighting by $E(C|A, B, X, Y)$ is equivalent to the estimating function in Equation (11). By iterated expectation,

$$\begin{aligned} & \sum_{i=1}^n E \left[E(C_i | A_i, B_i, X_i, Y_i) \{Y_i - \eta(A_i, X_i; \gamma)\} \frac{\partial \eta(A_i, X_i; \gamma)}{\partial \gamma^T} \right] \\ &= \sum_{i=1}^n E \left[C_i \{Y_i - \eta(A_i, X_i; \gamma)\} \frac{\partial \eta(A_i, X_i; \gamma)}{\partial \gamma^T} \right] = 0. \end{aligned}$$

Therefore, under suitable regularity conditions, we can obtain a consistent and asymptotically normal estimator for γ and $\mu(a, 0)$ provided that we can obtain a consistent estimator of $E(C|A, B, X, Y)$.

As discussed in Section 3.2, $E(C|A, B, X, Y)$ can be estimated directly from mixture components of the mixture density in (1). Thus we can consistently estimate $\mu(a, 0)$ by jointly solving Equations (2), (10), and (12). Because this estimator uses a weighted regression model, we refer to $\hat{\mu}(a, 0)$, given by the solution to these estimating equations, as the W-REG estimator. To be explicit, to obtain consistent estimation of $\mu(a, 0)$, W-REG requires correctly specifying each of the components of the mixture distribution in (1) (i.e., $\Pr(C|A = a, X = x, Y = y; a)$ and the distribution of $B|A, X, Y, C$) as well as the conditional mean model $\eta(A, X; \gamma)$.

Note that by estimating the parameters in Equation (3) we will already have estimated the regression coefficients of $\eta(A, X; \gamma)$. Thus, we can simply estimate $\mu(a, 0)$ by jointly solving the estimating Equations (4) and (10), which is equivalent to replacing $\eta(A = a, X; \gamma)$ in Equation (10) with $\eta(A = a, X; \hat{\gamma}_n)$. Because this estimator uses a regression model with coefficients estimated from the mixture distribution using an EM algorithm, we refer to this estimator as the EM-REG estimator. To be clear, EM-REG requires that the analyst correctly specify all the components of the mixture distribution in (3) (i.e., $\Pr(C = 1|A = a, X = x; a)$ and the distributions of $B|A, X, Y, C$ and $Y|A, X, C$).

W-REG has one potential advantage over EM-REG. The EM-REG estimator requires that the analyst specify a model for $\eta(A, X; \gamma)$ prior to fitting the mixture distribution, whereas W-REG first allows estimation of the biomarker mixture distribution without specifying a model for $\eta(A, X; \gamma)$. Because fitting mixture distributions can be computationally demanding, fitting several models for $\eta(A, X; \gamma)$ as part of the mixture distribution and assessing model fit may be challenging. In contrast, once the weights $E(C|A, X, B, Y; \hat{\xi}_n, \hat{\alpha}_n)$ for W-REG have been estimated, the analyst can easily fit multiple models and proceed using standard model checking and diagnostics for $\eta(A, X; \gamma)$.

3.4. Augmented Estimators

Augmented estimators incorporate both weighted estimators and regression-based estimators. In the case where C is observed, an augmented estimator for $\mu(a, 0)$ is the solution to the estimating equation

$$\sum_{i=1}^n I(A_i = a) \left[\frac{C_i Y_i}{E(C_i | A_i, X_i)} - \frac{\{C_i - E(C_i | A_i, X_i)\} \eta(A = a, X_i; \gamma)}{E(C_i | A_i, X_i)} - \mu(a, 0) \right] = 0. \quad (13)$$

This is similar to the IPCW estimator, but it is “augmented” with the term $\frac{\{C - E(C|A, X)\} \eta(A = a, X; \gamma)}{E(C|A, X)}$ which uses the regression function $\eta(A = a, X; \gamma)$. If the regression model $\eta(A = a, X; \gamma)$ is correctly specified, augmented estimators gain efficiency over IPCW estimators [3]. Additionally, assuming C is observed, the estimator for $\mu(a, 0)$ given by the solution to (13) is consistent when either the model for $E(C|A, X)$ is correctly specified or when the model for $\eta(A, X; \gamma)$ is correctly specified even if the other model is misspecified [6]. Thus, augmented estimators are said to be doubly-robust. In the case where C is unobserved, an augmented estimator for $\mu(a, 0)$ is the solution to the estimating equation

$$\sum_{i=1}^n I(A_i = a) \left[\frac{E(C_i | A_i, B_i, X_i, Y_i) Y_i}{E(C_i | A_i, X_i)} - \frac{\{E(C_i | A_i, B_i, X_i, Y_i) - E(C_i | A_i, X_i)\} \eta(A = a, X_i; \gamma)}{E(C_i | A_i, X_i)} - \mu(a, 0) \right] = 0.$$

This is similar to the estimating equation in (13), but we have replaced C , the indicator variable of compliance, with the conditional expectation $E(C|A, B, X, Y)$. Because this is an augmented CURE estimator, we refer to it as the A-CURE estimator.

For ease of exposition, we assume for the remainder of this section only a single treatment group with $A = 1$ for all participants, but the results easily generalize. The A-CURE estimator

$$\hat{\mu}_n(1, 0) = \frac{1}{n} \sum_{i=1}^n \frac{E(C_i | B_i, X_i, Y_i; \hat{\xi}_n, \hat{\alpha}_n) Y_i}{E(C_i | X_i; \hat{\beta}_n)} - \frac{\{E(C_i | B_i, X_i, Y_i; \hat{\xi}_n, \hat{\alpha}_n) - E(C_i | X_i; \hat{\beta}_n)\} \eta(A = 1, X_i; \hat{\gamma}_n)}{E(C_i | X_i; \hat{\beta}_n)}$$

is a consistent estimator for $\mu(1, 0)$ provided that the model for $E(C|B, X, Y)$ is correctly specified, and either the model for $E(C|X)$ or the model for $\eta(A = 1, X; \gamma)$ is correctly specified. That is, if the model for $E(C|B, X, Y)$ is correct, then $\hat{\mu}_n(1, 0)$ is a doubly-robust estimator for $\mu(1, 0)$.

To show this, we take the approach in [6] and use the conventions $\hat{\xi}_n \xrightarrow{P} \xi^*$, $\hat{\alpha}_n \xrightarrow{P} \alpha^*$, $\hat{\beta}_n \xrightarrow{P} \beta^*$, and $\hat{\gamma}_n \xrightarrow{P} \gamma^*$ to denote convergence in probability whether or not the models are correct (i.e., ξ^* is the “least wrong” parameter if the model is misspecified and similarly for α^* , β^* , and γ^*). If the model is correctly specified, then $\xi^* = \xi_0$, $\alpha^* = \alpha_0$, $\beta^* = \beta_0$, and $\gamma^* = \gamma_0$, that is, the subscript 0 indicates the true parameter values for the correct models. Because $\hat{\mu}_n(1, 0)$ is a sample mean, under suitable regularity conditions it converges in probability to

$$E\left[\frac{E(C|B, X, Y; \xi^*, \alpha^*)Y}{E(C|X; \beta^*)} - \frac{\{E(C|B, X, Y; \xi^*, \alpha^*) - E(C|X; \beta^*)\}\eta(A = 1, X; \gamma^*)}{E(C|X; \beta^*)}\right].$$

When $\xi^* = \xi_0$ and $\alpha^* = \alpha_0$, that is, when the models for the parameters of the mixture density (1) are correct, $\hat{\mu}_n(1, 0)$ converges in probability to

$$\begin{aligned} & E\left[\frac{E(C|B, X, Y; \xi_0, \alpha_0)Y}{E(C|X; \beta^*)} - \frac{\{E(C|B, X, Y; \xi_0, \alpha_0) - E(C|X; \beta^*)\}\eta(A = 1, X; \gamma^*)}{E(C|X; \beta^*)}\right] \quad (14) \\ &= E\left[\frac{CY}{E(C|X; \beta^*)} - \frac{\{C - E(C|X; \beta^*)\}\eta(A = 1, X; \gamma^*)}{E(C|X; \beta^*)}\right] \end{aligned}$$

The equality follows from iterated expectation. By the consistency assumption,

$$\frac{CY}{E(C|X; \beta^*)} = \frac{CY^*(1, 0)}{E(C|X; \beta^*)} = Y^*(1, 0) + \frac{\{C - E(C|X; \beta^*)\}Y^*(1, 0)}{E(C|X; \beta^*)}. \quad (15)$$

Substituting (15) into (14), we can write (14) as

$$E\{Y^*(1, 0)\} + E\left[\frac{\{C - E(C|X; \beta^*)\}\{Y^*(1, 0) - \eta(A = 1, X; \gamma^*)\}}{E(C|X; \beta^*)}\right]. \quad (16)$$

Because the first term in (16) is equal to $\mu(1, 0)$, $\hat{\mu}_n(1, 0)$ is doubly-robust provided that the second term in (16) equals 0. The second term equals 0 when either $\beta^* = \beta_0$ or $\gamma^* = \gamma_0$, that is, when either the model for $E(C|X; \beta)$ is correct or when the model for $\eta(A = 1, X; \gamma)$ is correct. As in [6], we will show this by taking iterated expectations.

First, assume that the model for $E(C|X, \beta)$ is correct, that is, $\beta^* = \beta_0$. By an iterated expectation conditioning on $Y^*(1, 0)$ and X , the second term in (16) equals

$$\begin{aligned}
& E \left(\frac{E\{C|X, Y^*(1, 0)\} - E(C|X; \beta_0)}{E(C|X; \beta_0)} \{Y^*(1, 0) - \eta(A = 1, X; \gamma^*)\} \right) \\
&= E \left(\frac{\{E(C|X) - E(C|X; \beta_0)\} \{Y^*(1, 0) - \eta(A = 1, X; \gamma^*)\}}{E(C|X; \beta_0)} \right) = 0,
\end{aligned}$$

where the equality follows from the assumption of no unmeasured confounding. Now assume that the model $\eta(A = 1, X; \gamma)$ is correctly specified, that is, $\gamma^* = \gamma_0$. By an iterated expectation conditioning on C and X , the second term in (16) equals

$$E \left[\frac{\{C - E(C|X; \beta^*)\} \{E\{Y^*(1, 0)|C, X\} - \eta(A = 1, X; \gamma_0)\}}{E(C|X; \beta^*)} \right]. \quad (17)$$

By the assumption of no unmeasured confounding, $E\{Y^*(1, 0)|C, X\} = E\{Y^*(1, 0)|C = 1, X\} = \eta(A = 1, X; \gamma_0)$. Hence, $\hat{\mu}_n(1, 0)$ is a doubly-robust estimator for $\mu(1, 0)$.

Note that an augmented estimator based on estimating $E(C|B, X, Y)$ using the mixture density (3) is less useful. In Equation (3), one must correctly specify a model for $Y|X, C$ *a priori* in order to consistently estimate $E(C|B, X, Y)$. That is, an augmented estimator based on estimating $E(C|B, X, Y)$ using components of the mixture density (3) would not be doubly robust.

For consistent estimation of $\mu(1, 0)$, the A-CURE and regression-based estimators *all* rely on the assumption that the distribution of $B|A, X, Y$ has been correctly specified. The regression-based estimators further require that $E(Y|A, X, C = 1)$ is correctly specified. In contrast, the A-CURE estimator is still consistent provided that *either* the mean model for $E(Y|A, X, C = 1)$ is correct *or* that the model for $E(C|A, X)$ is correct, essentially giving the analyst two chances to be correct instead of just one. This makes the A-CURE estimator attractive compared to the regression-based estimators.

3.5. Overview

In Table 1 we summarize the quantities that must be specified and estimated in order to achieve consistent and asymptotically normal estimators of $\mu(a, 0)$ for each estimator. We note that “correctly specified” in Table 1 refers to all components of the quantity (i.e. regression model, link function, distributional form, etc.) in order to achieve consistent and asymptotically normal estimators of $\mu(a, 0)$. We also indicate which components may be misspecified and still lead to consistent and asymptotically normal causal estimators. For comparison, we also include the quantities that must be correctly specified to achieve consistency in “traditional” IPCW, regression, and augmented IPCW estimators when the true compliance status is observed.

All estimators for $\mu(a, 0)$ will be consistent and asymptotically normal if the components in Table 1 are correctly specified. Let $\mathbf{Z}_i = (A_i, B_i, X_i, Y_i)$ be the observed data for the i th participant. Define $\psi_i(\mathbf{Z}_i; \mu(a, 0), \lambda)$ to be the multi-dimensional estimating function from

stacking the estimating functions for $\mu(a, 0)$ and any other nuisance parameters that must be estimated. For example, for the CURE estimator, λ would include $\lambda = (\xi^T, \alpha^T, \beta^T)^T$ and ψ would include Equations (6), (2), and (9). Under suitable regularity conditions,

$$\sqrt{n} \left[\left\{ \hat{\mu}(a, 0), \hat{\lambda}^T \right\}^T - \left\{ \mu(a, 0), \lambda^T \right\}^T \right] \xrightarrow{D} N \left(0, U^{-1} V (U^{-1})^T \right),$$

where $V = E[\psi_i(\mathbf{Z}_i; \mu(a, 0), \lambda)] [\psi_i(\mathbf{Z}_i; \mu(a, 0), \lambda)]^T$ and $U = -E \left[\frac{\partial \psi_i(\mathbf{Z}_i; \mu(a, 0), \lambda)}{\partial \{\mu(a, 0), \lambda^T\}^T} \right]$. The

sandwich covariance matrix $U^{-1} V (U^{-1})^T$ can be estimated using the empirical averages for U and V (see, for example, [14]) or with the nonparametric bootstrap [15].

4. Simulation

We conducted a simulation study to characterize the finite-sample performance of the estimators described in Section 3. For simplicity, we considered a scenario with only a single treatment group with $A=1$ for all participants. All estimators considered here require correctly specified models for the distribution of $B|A, X, Y, C$, and we do not investigate effects of misspecifying the conditional distribution of the biomarker. Instead we investigate the costs and benefits of modeling the distribution (or the conditional expectation) of $Y|A, X, C$ under correct or incorrect model specification.

To facilitate data generation, define the latent continuous measure of compliance C' . We generated (C', X_1, X_2) from a multivariate normal distribution with mean vector $(0, 2, 2)^T$ and covariance matrix $\begin{pmatrix} 1.000 & 0.200 & 0.075 \\ 0.200 & 1.000 & 0.200 \\ 0.075 & 0.200 & 1.000 \end{pmatrix}$. Using this covariance matrix, X_1 is more strongly associated with C' than is X_2 , and hence X_1 is more strongly associated with compliance. We defined the compliance indicator $C = I\{C' > \Phi^{-1}(0.8)\}$, where Φ is the standard normal distribution function. This gives $\Pr(C = 1) = 0.20$, roughly the estimated probability of compliance in the application we consider in Section 5. We simulated the outcome as $Y|X_1, X_2, C \sim N(5 - 2.25C + X_1 + X_2 + \nu X_3, 1.25^2)$, where $X_3 = X_1 X_2$ for a model with an interaction and $X_3 = X_1^2$ for a model with a quadratic term, and where ν was equal to 0, or 1.

For misspecified models, we omitted the term for X_3 when $\nu \neq 0$. Note that the CURE estimator is never misspecified in these scenarios because the estimator does not require estimation of the conditional distribution of Y . The biomarker B was generated as $B|Y, C$ from a normal distribution with mean equal to $-9.3 - 0.8C + 0.7Y$ and standard deviation equal to either 0.44, corresponding to an AUC of 0.9 (i.e. high AUC), or 0.67, corresponding to an AUC of 0.8 (i.e. low AUC). Our data generating mechanism implies that B and X are conditionally independent given C and Y . Although this assumption is not required, it simplifies the simulation scenario, because in the correctly specified conditional mean function of B in the mixture densities in (1) and (3), coefficients for X_1 and X_2 are 0 and thus do not need to be estimated. As in Boatman et al. [10], we included participants comprising 10% of the sample who were known to be compliant, that is, $C = 1$ without error. Data for known compliers was generated as described above, but, rather than randomly

generating C , we set $C = 1$ for these subjects. These known compliers contributed to the estimation of the mixture density as described in Section 3. We considered sample sizes of 275, 550, and 1,100, including the known compliers. 1,000 Monte Carlo datasets were generated for each scenario.

We compared eight estimators of $\hat{\mu}(1, 0)$: the five estimators described in Section 3 and three traditional estimators that use the true compliance status C : 1) an IPCW estimator given by the solution to Equation (5), 2) a regression-based estimator (REG) given by solution to Equation (10) that estimates the linear model $Y|X, C=1$ using least-squares and 3) an augmented IPW estimator (A-IPW) given by solution to Equation (13) that combines the IPW and REG estimators. Although C would generally not be available to the data analyst, we included these three estimators to illustrate the consequences of relying on imperfect measures of compliance.

For all models that require estimation of a mixture distribution to estimate $E(C|B, X, Y)$, we assumed a linear regression model with normally distributed errors for $B|Y, X, C$. For the CURE estimator, we estimated $E(C|X, Y)$ using a generalized linear model (GLM) with logit link. Note that for the CURE estimator, this GLM is not exactly concordant with the data generating mechanism, but it is simple, easy to implement, and achieves reasonable small-sample performance. For the CURE+ estimator, we assumed a linear regression model for $Y|X, C$ with normally distributed errors, and we estimated $E(C|X)$ using a GLM with probit link. For the EM-REG estimator, we used the estimated model for $Y|X, C$ to solve the estimating equation given in Equation (10). For the W-REG estimator, we used the estimated expectation $E(C|B, X, Y)$ to solve Equation (12) and used the coefficient estimates to solve Equation (10). For the A-CURE estimator, we obtained estimates of the necessary components using the same approach as in the CURE and W-REG estimators. 95% Confidence intervals for all estimators were based on a normal approximation with standard error estimated by the non-parametric bootstrap with 100 bootstrap samples.

For each model requiring estimation via the EM algorithm, we tried 4 sets of starting values, and we chose the model that gave the lowest negative log likelihood. The first set of starting values used parameters and coefficients estimated from models using the actual compliance C . The second set of starting values were the true values of the parameters and coefficients; for misspecified models, these were found by estimating parameters and coefficients from misspecified models using a very large sample. The remaining sets of starting values were generated from the first set of starting values as follows. For a coefficient or parameter D from the first set, the new starting value was $D + D \times U$, where U is a uniform random variable on the interval $(-0.35, 0.35)$.

Scenarios in which the model for $Y|X, C$ is misspecified resulted in some estimates of $\mu(1, 0)$ which were clearly outliers in the sampling distribution. In practice, investigators are likely to fit a variety of causal models to seek converging lines of evidence of the magnitude of a causal effect. Thus, outliers are generally easily identified and omitted in practice, and we feel interpretation of simulation results may be more meaningful with outliers removed. We present results with outliers excluded, and we show results with outliers included in the Supplementary Materials. To remove outliers, for each estimator within each scenario, we

defined upper and lower limits as the median $\pm 9/4 \times \text{IQR}$ (interquartile range; $9/4 \times \text{IQR}$ is roughly 3 standard deviations for a normally distributed random variable).

Table 2 and the top halves of Tables 3, and 4 provide results when the regression model for the outcome given confounders X and compliance C is correctly specified. All estimators show a small amount of bias that is attenuated with increasing sample size. Among the “gold standard” estimators which use the true compliance status, all were more efficient than the proposed estimators that do not use the true compliance status. Across the scenarios presented in tables 2, 3, and 4, the “gold standard” regression estimator was substantially more efficient than the IPW estimator (the ratio of REG MSE to IPW MSE ranged from 0.775 to 0.924). However, as expected, the augmented IPW estimator recovers nearly all the difference in efficiency between IPW and REG.

Considering the estimators that do not use the true compliance status, CURE+ was much more efficient than CURE when the true outcome model was correctly specified. Using more data (i.e., the outcome) may more efficiently identify the components of the mixture distribution leading to more efficient estimation of $E(C|A, B, X, Y)$. As is the case for “gold standard” estimators, regression-based estimators were generally more efficient when the outcome model is correctly specified. EM-REG, which uses a maximum likelihood estimator for the parameters of the outcome model, was more efficient than W-REG. Augmenting the CURE estimator (i.e., A-CURE) improves the efficiency relative to the CURE estimator with MSEs generally comparable to the W-REG estimator. Coverage of the 95% percent non-parametric bootstrap confidence intervals resulted in over-coverage with $n = 275$ but were closer to the nominal level in larger sample sizes. All estimators exhibited worse performance when the AUC for the biomarker was 0.8, but better performance was observed with larger sample sizes.

The bottom halves of Tables 3 and 4 show the results when the coefficient for the interaction and quadratic terms are non-zero but the model is misspecified, that is, the regression model for the outcome omits the interaction or quadratic term. For the interaction model, CURE performs well because there is no model misspecification. CURE+, EM-REG and W-REG show bias that is only weakly attenuated with increasing sample size. For the quadratic models with misspecification, the bias of the CURE+, EM-REG, and W-REG estimators is more substantial than for the interaction. The differences in the performance of these estimators is explained by noting that the quadratic models include a quadratic term for X_1 , and, because X_1 is more strongly associated with compliance than X_2 , the models with misspecification of $E(Y|X, C=1)$ tend to perform worse under the quadratic model than under the interaction model. The A-CURE estimator has bias commensurate with the CURE estimator, but the MSE of the A-CURE estimator is now higher than the MSE of the CURE estimator. This is consistent with the performance of the “gold standard” estimators in which the AIPW estimator was less efficient than the IPW estimator when the outcome regression model was misspecified. Coverage of the 95% percent non-parametric bootstrap confidence intervals resulted in Monte Carlo coverage probabilities greater than 0.95 for $n = 275$ (when over-coverage was observed in the correctly specified scenarios) but the coverage decreases for CURE+, EM-REG and W-REG as the sample size increases, whereas the CURE and A-CURE estimators achieve correct coverage for all samples sizes.

5. Application to CENIC-p1 Data

We applied the five estimators described in Section 3 to data from CENIC-p1 [7] to estimate the causal effect of VLNC cigarettes on the number of cigarettes smoked per day during week 6 of the trial. For comparison, we also included the ITT estimator and a per protocol estimator based on self-reported compliance. CENIC-p1 included seven treatment groups: a usual brand control condition and six experimental conditions with nicotine content ranging from 15.8 mg/g of tobacco to 0.4 mg/g of tobacco. In addition, one condition had 0.4 mg/g of tobacco and high tar to evaluate the impact of tar yield on the effect of nicotine reduction. For the purposes of this illustration, we focus on the contrast between the two VLNC conditions (i.e. 0.4 mg/g and the 0.4 mg/g, high tar conditions, combined) versus the usual brand control condition. As in Boatman et al [10], we only considered compliance and outcomes during week 6 of the trial. We let $A = 1$ if the participant was assigned to smoke VLNC cigarettes ($n = 242$), and $A = 2$ if the participant was assigned to smoke usual brand cigarettes ($n = 118$). In this study, subjects are considered noncompliant if they smoke cigarettes which are not provided as part of the study (i.e., commercial or “non-study” cigarettes).

The goal of the analysis was to estimate $\mu(2, 0) - \mu(1, 0)$, the expected reduction in cigarettes smoked per day during week 6 if smoking only VLNC cigarettes. Because the usual brand group is meant to represent smoking of commercial cigarettes, all participants in this group were treated as compliant even if they reported smoking non-study cigarettes (i.e., cigarettes that were not provided by trial personnel). Thus, $\mu(2, 0)$ is the sample average of the number of cigarettes smoked per day for this group.

For the CURE and CURE+ estimators, we estimated the probability of compliance following the approach in Section 3.2 using the (natural) logarithm of total nicotine equivalent (TNE) measured at week 6 as the biomarker B . In fitting the mixture distributions in Equations (1) and (3), we assumed a linear regression models for $B|Y, X, C$ with normally distributed errors and no shared parameters between different levels of C . As in the simulation, we assumed that B and X are conditionally independent given C and Y . As described in [16], CENIC-p1 also included an auxiliary study that included data from 23 smokers who volunteered to be sequestered in a hotel for 4 nights with access to only VLNC cigarettes. These participants are known to be compliant, and we incorporated B and Y from these individuals to assist in estimation of the mixture densities. These individuals contribute data to the likelihood for the distribution of $B|Y, C = 1$, but they make no other contributions to estimation (i.e., their outcomes are not used to estimate the causal effect, and their covariates X are not used in any way). We assumed logistic regression models for $\Pr(C = 1|X, Y)$ and $\Pr(C = 1|X)$ of Equations (1) and (3), where the confounders X included age, level of addiction (baseline cigarettes per day and log of TNE), measures of withdrawal (Minnesota Nicotine Withdrawal Scale at week 5 and maximum acute withdrawal), and satisfaction with and craving for VLNC (Cigarette Evaluation Scale and Questionnaire of Smoking Urges at week 5) and normal nicotine cigarettes (Questionnaire of Smoking Urges at Week 5). For the distribution of $Y|X, C$, we assumed linear regression models with linear terms for all X , with normally distributed errors and no shared parameters between different levels of C . We assumed that individuals with $A = 1$ who self-reported any noncompliance

(that is, $D = 0$) reported noncompliance without error, and we let $\Pr(C = 1|X, Y) = 0$ for these individuals. As a result, these individuals did not contribute to estimation of the mixture distribution. For both the CURE and CURE+ estimators, we assumed a logistic model for $E(C|X)$ using X as described above. For the EM-REG estimator, we used the estimated model for $Y|X, C$ from the mixture model (3) described above to solve the estimating equation given in Equation (10). For the W-REG estimator, we used the estimated expectation $E(C|B, X, Y)$ from the mixture model (1) to solve Equation (12) and used the coefficient estimates to solve Equation (10). For the A-CURE estimator, we combined the CURE and W-REG estimators. We used the bootstrap percentile method with 1,000 bootstrap re-sampled data sets to compute 95% confidence intervals. In the Supplementary Materials we show point estimates and 95% CIs for all all components required to implement the various causal estimators.

Table 5 shows the estimated causal effect of VLNC cigarettes on number of cigarettes smoked per day. The CURE estimator gives the most optimistic estimate of the causal effect, followed by the per protocol estimator, which only considers subjects that self-reported compliance to VLNC cigarettes (i.e. subjects that self-reported no non-study cigarette use). The W-REG and A-CURE estimators give similar estimates and are close to the ITT estimate. The CURE+ estimator and EM-REG estimators give similar and more conservative estimates. As expected from the simulation, the A-CURE estimator and the regression-based estimators have smaller estimated standard error (SE) than the CURE estimator. Importantly, all the causal estimators considered here showed a similar, statistically significant reduction in average cigarette per day consumption for those subjects smoking VLNC cigarettes.

In most therapeutic trials we would expect the causal estimates to be less conservative than the ITT estimates as noncompliance with the medication should dilute the treatment effect by reducing the dose received. However, the fact that the causal estimates in this application bracket the ITT estimate is not surprising because the impact of non-compliance may be due to two different factors which have different implications for the causal effect. First, non-compliant use of high nicotine cigarettes could more effectively alleviate withdrawal symptoms than study cigarettes with reduced nicotine. Consequently, one might expect that in order to achieve the same nicotine consumption, the number of study cigarettes smoked per day would be much higher if such participants were forced to be compliant. Second, non-compliance is also associated with individuals who find VLNC cigarettes particularly unsatisfying. Such individuals might smoke less if forced to be compliant. The results from this analysis provide some preliminary evidence that these two factors are relatively equal and may offset each other.

Although reductions in bias may be possible by including interaction and higher-order terms, the limited sample size in this application restricts our ability to explore more complex models for the response, Y . Consequently, we do not recommend the CURE+ and EM-REG estimators in practice unless a large sample size is available. Although the W-REG estimator has smaller SE than the CURE+ and EM-REG estimators, for the same reason we do not recommend this estimator for small samples. Finally, we note that the estimators presented here require us to assume that we have correctly modeled a variety of conditional densities described in Table 1 and should be interpreted cautiously. As for all models, the

assumptions must be considered in the interpretation, and converging lines of evidence should be sought to clarify causal mechanisms.

6. Discussion

Although a variety of well-established methods are available to estimate causal effects from RCTs, these methods assume that compliance is measured without error. This is an unreasonable assumption in many cases, particularly when compliance is measured by participants' self-report. We explored a variety of novel weighted estimators, regression-based estimators, and an augmented estimator for use when compliance is measured with error. Rather than assuming a measurement error model where self-reported compliance is assumed to be a surrogate for actual compliance, we treated compliance as an unobserved variable. We assumed that a biomarker of exposure is measured which can discriminate between compliers and non-compliers. In many pharmaceutical trials, the biomarker might be the concentration of drug metabolite in the blood or urine.

All the proposed models performed well when the models were correctly specified. As is the case when the compliance status is known, the regression-based estimators (EM-REG and W-REG) are most efficient, likely because they are able to average over all participants rather than assigning very large weights to a few participants (leading to estimators which are heavily influenced by the outcomes of a few individuals and, therefore, estimators with increased variability). However, the regression-based estimators and the CURE+ estimator do not perform well when the mean model for $Y|A, X, C=1$ is misspecified because they rely on correctly specifying the distribution or mean model for $Y|A, X, C=1$ for consistency. Furthermore, the model for $Y|A, X, C=1$ is fit as part of a mixture distribution for the EM-REG estimator. Fitting mixture distributions can be computationally challenging which may preclude fitting several models for $E(Y|A, X, C=1)$ and assessing model fit. Although the W-REG estimator allows the analyst to fit multiple regression model for $Y|A, X, C=1$ without refitting a mixture distribution, the efficiency gains of the W-REG estimator were more modest and was not substantially better than the A-CURE estimator. For these reasons, we do not recommend the EM-REG or the W-REG estimators in practice unless investigators have confidence that the model can be at least approximately correctly specified.

Additionally, when the outcome model is substantially misspecified, it is possible that the maximum likelihood estimates of a two-component mixture model will not correspond to the compliers and noncompliers. Instead the mixture components can be used to better approximate the misspecification in the outcome model. In this case, the CURE+ and EM-REG estimators are substantially biased. This emphasizes our conclusion that methods such as CURE+ and EM-REG, which require correct specification of the outcome model as part of fitting the mixture distribution, are less preferable to doubly robust approaches like A-CURE.

Although we did not explore misspecification of the models for the CURE estimator, the CURE estimator has intuitive appeal because it relies primarily on the mixture distribution of the biomarker to discriminate compliers and non-compliers. As we demonstrated via

simulation, the A-CURE estimator can be more efficient than the CURE estimator by augmenting the IPCW-like CURE estimator with a regression model for the response given confounders and compliance status. The A-CURE estimator has the appealing property of being doubly-robust in the sense that either model for $Y|A, X, C = 1$ or the model for $C|A, X$ must be correctly specified in order for the estimator to be consistent and asymptotically normal. But we note that the mixture distribution in Equation (1) must be correctly specified regardless for consistency.

The current work has some important limitations. As for all causal methods, the estimators require strong and untestable assumptions, and the causal effects estimated in the application must be interpreted with this in mind. Furthermore, the methods developed here assume that compliance is measured at a specific time point and do not incorporate information on the amount of self-reported noncompliance. For example, in the application, we considered outcomes and compliance only during week 6 of the trial and did not incorporate information on the number of non-study cigarettes smoked. A natural extension of the proposed methods developed here would incorporate longitudinal data and information on the (self-reported) level of noncompliance. Finally, while our simulation study only considers a two-dimension covariate vector, X , in order to focus our investigation on the robustness vs. efficiency trade-off associated with the various proposed estimators, practical application of our methods would likely involve higher-dimensional covariates with uncertainty regarding the covariates that must be included in order to satisfy the “no unmeasured confounders” assumption. Variable selection in the context of causal inference has been the focus of recent methodological development [17], which could form the basis of an extension of our method to the higher-dimensional setting but would likely come at the cost of efficiency. This is a likely subject of future work.

In Section 3.2, we showed how the components of the mixture densities presented in Equations 1 and 3 can be used to derive the probability that a subject was compliant conditional on their biomarker value when self-reported compliance is not reliable. In this paper, we proposed extensions of the IPCW, regression and augmented estimators presented in [3] that account for uncertainty in a subject’s compliance status when a biomarker must be used in place of self-reported non-compliance. Inaccurate self-reporting of compliance also poses challenges in other causal frameworks (e.g. principal stratification [18]), as well, and our approach may also provide a path for addressing uncertainty in compliance status in other frameworks but the technical details of how this would be implemented requires further investigation.

The estimators presented in Section 3 require that we correctly specify a mixture distribution for the biomarkers in order to obtain weights when self-reported compliance is not reliable. Previous research has shown that inverse probability weighted estimators for continuous treatments that rely on a continuous density estimate may be sensitive to model misspecification [19]. As a result, our proposed estimators are presumably more sensitive to model misspecification than standard IPCW estimators. However, we are only considering the case of binary treatment (i.e. compliant vs. non-compliant) and, therefore, expect that our estimators would be more robust to model misspecification than inverse probability weighted estimators for continuous treatment.

We developed and evaluated the performance of a series of novel estimators of causal effects when compliance is measured with error. The simulation and application results suggest that the CURE and A-CURE estimators are good choices for investigators, but efficiency gains can be made by using the regression-based estimators if investigators are willing to assume that regression model $Y|A, X, C = 1$ can be specified approximately correctly.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Eric C. Donny for helpful comments on the methods and application.

9. Funding

This research was partially funded by NIH grant P30-CA077598 and R01-CA214824 from the National Cancer Institute and R03-DA041870 and U54-DA031659 from the National Institute on Drug Abuse and FDA Center for Tobacco Products (CTP). The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH or FDA CTP.

References

- Hernán MA, Hernández-Díaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clinical Trials* 2012; 9:48–55. [PubMed: 21948059]
- Bellamy SL, Lin JY, Ten Have TR. An introduction to causal modeling in clinical trials. *Clinical Trials* 2007; 4:58–73. [PubMed: 17327246]
- Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Statistics in Medicine* 2004; 23:2937–2960. [PubMed: 15351954]
- Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health* 2006; 60(7):578–586. [PubMed: 16790829]
- Robins JM, Hernán MA. Longitudinal Data Analysis, chap. 23: Estimation of the Causal Effects of Time-Varying Exposures Boca Raton: Chapman & Hall/CRC, 2008.
- Tsiatis AA. Semiparametric Theory and Missing Data. Springer Science+Business Media, LLC, 2006.
- Donny EC, Denlinger RL, Tidey JW, Koopmeiners JS, Benowitz NL, Vandrey RG, al' Absi M, Carmella SG, Cinciripini PM, Dermody SS, et al. Randomized trial of reduced-nicotine standards for cigarettes. *The New England Journal of Medicine* 2015; 373:1340–9. [PubMed: 26422724]
- Valeri L, Lin X, VanderWeele TJ. Mediation analysis when a continuous mediator is measured with error and the outcome follows a generalized linear model. *Statistics in Medicine* 2014; 33:4875–4890. [PubMed: 25220625]
- Ogburn EL, VanderWeele TJ. Analytic results on the bias due to nondifferential misclassification of a binary mediator. *American Journal of Epidemiology* 2012; 176(6):555–561. [PubMed: 22930481]
- Boatman JA, Vock DM, Koopmeiners JS. Estimating causal effects from a randomized clinical trial when noncompliance is measured with error. *Biostatistics* 2018; 19(1):103–118. [PubMed: 28605411]
- Cole SR, Hernán MA. Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology* 2008; 168(6):656–664. [PubMed: 18682488]
- Direct Pearl J. and indirect effects. *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence*; San Francisco: Morgan Kaufmann; 2001. 411–420.
- Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society Series B-Methodological* 1977; 39:1–38.

14. Stefanski LA, Boos DD. The calculus of m-estimation. *The American Statistician* 2002; 56(1):29–38.
15. Efron B Bootstrap methods: Another look at the jackknife. *The Annals of Statistics* 1979; 7(1):1–26.
16. Denlinger RL, Smith TT, Murphy SE, Koopmeiners JS, Benowitz NL, Hatsukami DK, Pacek LR, Colino C, Cwalina SN, Donny EC. Nicotine and anatabine exposure from very low nicotine content cigarettes. *Tobacco Regulatory Science* 2016; 2(2):186–203. [PubMed: 29600258]
17. Koch B, Vock DM, Wolfson J. Covariate selection with group lasso and doubly robust estimation of causal effects. *Biometrics*; :n/a–n/a.
18. Frangakis CE, Rubin DB. Principal stratification in causal inference. *Biometrics* 2002; 58(1):21–29. [PubMed: 11890317]
19. Hernán MA, Robins JM. *Causal Inference* Chapman & Hall/CRC, 2018.

Table 1.

Summary of components which must be estimated for each of the five proposed causal estimators. A checkmark indicates the component must be estimated. A dash indicates the component is not estimated. The last column indicates which terms may be misspecified and still have the causal estimator to be consistently and asymptotically normal (CAN).

Estimator	$f(B A, X, Y, C)$	$E(C A, X)$	$E(C A, X, Y)$	$h(Y A, X, C)$	$E(Y A, X, C)$	Terms Misspecified but Estimator CAN
IPW	–	✓	–	–	–	None
REG	–	–	–	–	✓	None
A-IPW	–	✓	–	–	✓	$E(C A, X)$ or $E(Y A, X, C)$
CURE	✓	✓	✓	–	–	None
CURE+	✓	✓	–	✓	✓	None
EM-REG	✓	✓	–	✓	✓	None
W-REG	✓	–	✓	–	✓	None
A-CURE	✓	✓	✓	–	✓	$E(C A, X)$ or $E(Y A, X, C)$

Table 2.

Simulation results for the estimators discussed in Section 3 when the true outcome model does not include an interaction or quadratic term and the outcome model is correctly specified. Results are presented assuming both high AUC (AUC = 0.9) and low AUC (AUC = 0.8). Presented are the simulated bias, Monte Carlo standard deviation (MCSD), mean squared error (MSE), coverage probability of 95% confidence interval (Cov), and number of outliers (n_O).

Estimator	$n = 275$					$n = 550$					$n = 1100$				
	Bias	MCSD	MSE	Cov	n_O	Bias	MCSD	MSE	Cov	n_O	Bias	MCSD	MSE	Cov	n_O
High AUC															
IPW	0.013	0.230	0.053	0.947	2	0.014	0.158	0.025	0.948	4	0.003	0.110	0.012	0.954	4
REG	0.006	0.215	0.046	0.947	1	0.009	0.145	0.021	0.941	4	-0.001	0.102	0.010	0.943	3
A-IPW	0.007	0.215	0.046	0.943	2	0.010	0.147	0.022	0.940	4	-0.001	0.102	0.010	0.945	4
CURE	-0.044	0.387	0.152	0.975	8	-0.011	0.263	0.069	0.972	1	-0.019	0.179	0.033	0.954	0
CURE+	0.005	0.304	0.092	0.961	6	0.006	0.211	0.045	0.957	0	-0.006	0.153	0.023	0.949	0
EM-REG	-0.008	0.302	0.091	0.965	3	0.000	0.203	0.041	0.958	1	-0.007	0.145	0.021	0.946	0
W-REG	-0.062	0.401	0.165	0.970	6	-0.013	0.258	0.067	0.971	2	-0.018	0.174	0.031	0.951	0
A-CURE	-0.061	0.403	0.166	0.968	4	-0.011	0.259	0.067	0.967	2	-0.018	0.175	0.031	0.951	0
Low AUC															
IPW	0.013	0.230	0.053	0.947	2	0.014	0.158	0.025	0.948	4	0.003	0.110	0.012	0.954	4
REG	0.006	0.215	0.046	0.947	1	0.009	0.145	0.021	0.941	4	-0.001	0.102	0.010	0.943	3
A-IPW	0.007	0.215	0.046	0.943	2	0.010	0.147	0.022	0.940	4	-0.001	0.102	0.010	0.945	4
CURE	-0.143	0.563	0.337	0.992	90	-0.054	0.406	0.167	0.986	67	-0.041	0.281	0.081	0.971	28
CURE+	-0.020	0.346	0.120	0.980	39	-0.001	0.245	0.060	0.976	10	-0.010	0.174	0.030	0.957	9
EM-REG	-0.034	0.351	0.124	0.978	35	-0.010	0.240	0.058	0.974	11	-0.013	0.169	0.029	0.963	6
W-REG	-0.178	0.603	0.395	0.989	81	-0.071	0.421	0.182	0.983	55	-0.040	0.277	0.078	0.971	28
A-CURE	-0.181	0.603	0.396	0.990	83	-0.071	0.421	0.182	0.982	56	-0.039	0.278	0.079	0.969	29

Table 3.

Simulation results for the estimators discussed in Section 3 when the true outcome model contains an interaction term. Results are presented with the model is correctly specified and under model misspecification. Presented are the simulated bias, Monte Carlo standard deviation (MCSD), mean squared error (MSE), coverage probability of 95% confidence interval (Cov), and number of outliers (n_O).

Estimator	$n = 275$					$n = 550$					$n = 1100$				
	Bias	MCSD	MSE	Cov	n_O	Bias	MCSD	MSE	Cov	n_O	Bias	MCSD	MSE	Cov	n_O
Model Correctly Specified															
IPW	0.028	0.393	0.155	0.962	1	0.019	0.273	0.075	0.959	2	0.007	0.187	0.035	0.962	2
REG	0.016	0.350	0.122	0.950	4	0.013	0.245	0.060	0.945	2	0.002	0.171	0.029	0.954	1
A-IPW	0.016	0.350	0.123	0.949	4	0.013	0.246	0.061	0.944	2	0.002	0.172	0.029	0.956	1
CURE	-0.029	0.527	0.278	0.993	25	-0.014	0.349	0.122	0.984	5	-0.018	0.236	0.056	0.959	4
CURE+	0.020	0.448	0.201	0.966	7	0.013	0.299	0.090	0.959	3	-0.003	0.214	0.046	0.952	0
EM-REG	-0.001	0.412	0.170	0.965	5	0.005	0.278	0.077	0.959	2	-0.005	0.201	0.040	0.955	1
W-REG	-0.067	0.488	0.242	0.973	7	-0.015	0.318	0.101	0.968	5	-0.019	0.225	0.051	0.951	0
A-CURE	-0.063	0.487	0.241	0.982	12	-0.015	0.319	0.102	0.969	4	-0.019	0.225	0.051	0.954	0
Model Misspecified															
IPW	0.028	0.393	0.155	0.962	1	0.019	0.273	0.075	0.959	2	0.007	0.187	0.035	0.962	2
REG	-0.031	0.377	0.143	0.944	3	-0.033	0.265	0.071	0.942	1	-0.042	0.182	0.035	0.957	3
A-IPW	0.004	0.384	0.147	0.944	5	0.005	0.269	0.072	0.947	2	-0.002	0.185	0.034	0.962	3
CURE	-0.029	0.527	0.278	0.993	25	-0.014	0.349	0.122	0.984	5	-0.018	0.236	0.056	0.959	4
CURE+	0.127	0.608	0.385	0.971	29	0.206	0.364	0.175	0.952	13	0.206	0.253	0.106	0.868	4
EM-REG	-0.043	0.574	0.331	0.974	43	0.048	0.343	0.120	0.979	15	0.048	0.244	0.062	0.950	1
W-REG	-0.095	0.538	0.298	0.973	10	-0.058	0.359	0.132	0.967	4	-0.059	0.248	0.065	0.958	2
A-CURE	-0.047	0.565	0.321	0.989	16	-0.007	0.361	0.130	0.979	7	-0.015	0.248	0.061	0.965	1

Table 4.

Simulation results for the estimators discussed in Section 3 when the true outcome model contains a quadratic term. Results are presented with the model is correctly specified and under model misspecification. Presented are the simulated bias, Monte Carlo standard deviation (MCSD), mean squared error (MSE), coverage probability of 95% confidence interval (Cov), and number of outliers (n_O).

Estimator	$n = 275$					$n = 550$					$n = 1100$				
	Bias	MCSD	MSE	Cov	n_O	Bias	MCSD	MSE	Cov	n_O	Bias	MCSD	MSE	Cov	n_O
Model Correctly Specified															
IPW	0.019	0.421	0.178	0.958	0	0.020	0.288	0.083	0.950	1	0.000	0.196	0.038	0.954	4
REG	0.011	0.386	0.149	0.952	4	0.016	0.272	0.074	0.946	3	-0.001	0.188	0.035	0.954	3
A-IPW	0.012	0.385	0.149	0.953	5	0.016	0.273	0.075	0.944	2	-0.001	0.188	0.035	0.956	3
CURE	-0.031	0.570	0.326	0.996	35	-0.000	0.365	0.133	0.990	10	-0.019	0.250	0.063	0.971	6
CURE+	0.015	0.484	0.234	0.967	3	0.012	0.313	0.098	0.962	3	-0.008	0.223	0.050	0.959	5
EM-REG	-0.001	0.446	0.199	0.963	4	0.009	0.305	0.093	0.956	2	-0.008	0.215	0.046	0.956	3
W-REG	-0.081	0.522	0.279	0.978	11	-0.021	0.340	0.116	0.972	7	-0.025	0.238	0.057	0.958	2
A-CURE	-0.082	0.526	0.283	0.984	12	-0.021	0.342	0.117	0.973	6	-0.025	0.238	0.057	0.958	2
Model Misspecified															
IPW	0.019	0.421	0.178	0.958	0	0.020	0.288	0.083	0.950	1	0.000	0.196	0.038	0.954	4
REG	-0.121	0.448	0.215	0.925	1	-0.109	0.320	0.114	0.926	1	-0.121	0.216	0.061	0.917	3
A-IPW	-0.018	0.456	0.208	0.942	6	0.002	0.338	0.114	0.941	1	-0.014	0.222	0.050	0.946	4
CURE	-0.031	0.570	0.326	0.996	35	-0.000	0.365	0.133	0.990	10	-0.019	0.250	0.063	0.971	6
CURE+	0.459	0.534	0.496	0.969	30	0.484	0.357	0.362	0.891	14	0.468	0.254	0.283	0.649	4
EM-REG	0.260	0.510	0.327	0.985	37	0.294	0.350	0.208	0.961	16	0.269	0.243	0.132	0.864	4
W-REG	-0.226	0.646	0.468	0.957	10	-0.141	0.418	0.195	0.963	6	-0.142	0.291	0.105	0.932	1
A-CURE	-0.078	0.675	0.461	0.984	25	-0.009	0.434	0.189	0.978	7	-0.021	0.295	0.088	0.965	2

Table 5.

Results of applying the estimators discussed in Section 3 to the data from CENIC-p1. Presented are the point estimates, standard error of the estimators (SE), and 95% confidence interval of the estimated causal effect for each estimator.

Estimator	$\hat{\mu}(2, 0)$	$\hat{\mu}(1, 0)$	$\hat{\mu}(2, 0) - \hat{\mu}(1, 0)$	SE	95% CI
ITT	22.18	15.37	6.81	1.57	(3.78, 10.07)
Per Protocol	22.18	15.12	7.07	1.80	(3.61, 10.63)
Causal Estimators					
CURE	22.18	14.98	7.20	2.87	(1.46, 11.63)
CURE+	22.18	16.98	5.20	3.11	(0.36, 11.97)
EM-REG	22.18	16.80	5.38	2.73	(1.56, 12.54)
W-REG	22.18	15.53	6.65	2.48	(1.85, 11.58)
A-CURE	22.18	15.47	6.71	2.55	(1.75, 11.48)

$\mu(2, 0)$: mean cigarettes smoked per day for the usual brand group.

$\mu(1, 0)$: mean cigarettes smoked per day for the VLNC group if all participants were to be compliant.