

Estimating causal effects from a randomized clinical trial when noncompliance is measured with error

JEFFREY A. BOATMAN*, DAVID M. VOCK, JOSEPH S. KOOPMEINERS

*Division of Biostatistics, University of Minnesota, A460 Mayo Building, MMC 303 420 Delaware St. SE,
Minneapolis, MN 55455, USA*

jeffrey.boatman@gmail.com

ERIC C. DONNY

*Department of Psychology, University of Pittsburgh, 4119 Sennott Square 210 S. Bouquet St.,
Pittsburgh, PA 15260, USA*

SUMMARY

Noncompliance or non-adherence to randomized treatment is a common challenge when interpreting data from randomized clinical trials. The effect of an intervention if all participants were forced to comply with the assigned treatment (i.e., the causal effect) is often of primary scientific interest. For example, in trials of very low nicotine content (VLNC) cigarettes, policymakers are interested in their effect on smoking behavior if their use were to be compelled by regulation. A variety of statistical methods to estimate the causal effect of an intervention have been proposed, but these methods, including inverse probability of compliance weighted (IPCW) estimators, assume that participants' compliance statuses are reported without error. This is an untenable assumption when compliance is based on self-report. Biomarkers (e.g., nicotine levels in the urine) may provide more reliable indicators of compliance but cannot perfectly discriminate between compliers and non-compliers. However, by modeling the distribution of the biomarker as a mixture distribution and writing the probability of compliance as a function of the mixture components, we show how the probability of compliance can be directly estimated from the data even when compliance status is unknown. To estimate the causal effect, we develop a new approach which re-weights participants by the product of their probability of compliance given the observed data and the inverse probability of compliance given confounders. We show that our proposed estimator is consistent and asymptotically normal and show that in some situations the proposed approach is more efficient than standard IPCW estimators. We demonstrate via simulation that the proposed estimator achieves smaller bias and greater efficiency than *ad hoc* approaches to estimating the causal effect when compliance is measured with error. We apply our method to data from a recently completed randomized trial of VLNC cigarettes.

Keywords: Causal inference; Clinical trials; Inverse probability weighting; Noncompliance; Regulatory science; Very low nicotine content cigarettes.

*To whom correspondence should be addressed.

1. INTRODUCTION

Twenty years ago, [Benowitz and Henningfield \(1994\)](#) argued the addictive properties of cigarettes could be eliminated if the nicotine content were reduced to 0.4–0.5 milligrams (mg) per gram of tobacco. In the United States, the Family Smoking Prevention and Tobacco Control Act provides the Food and Drug Administration (FDA) with the regulatory authority to limit the nicotine content of cigarettes to lower levels (but not zero) if such a regulation is likely to improve public health. As smoking remains the United States' leading cause of preventable death ([U.S. Department of Health and Human Services, 2014](#)), nicotine reduction could have a substantial public health impact. However, evidence for the effectiveness of such a policy is limited.

We recently reported the results of The Center for The Evaluation of Nicotine in Cigarettes, project 1 (CENIC-p1), a 6-week randomized trial evaluating the effect of nicotine reduction on tobacco use and dependence ([Donny and others, 2015](#)). Current smokers ($n = 839$) were randomized equally to one of seven groups consisting of a usual brand control condition or experimental cigarettes with nicotine content ranging from 15.8 mg per gram of tobacco (normal nicotine controls) to 0.4 mg per gram of tobacco. In addition, the investigators included a group that received cigarettes with 0.4 mg of nicotine per gram of tobacco with high tar to understand the effect of tar yield on cigarette use and dependence when nicotine content is reduced. Participants were instructed to smoke only those cigarettes provided in the trial and were considered non-compliant if they smoked cigarettes not provided by the trial (i.e., non-study commercial cigarettes). Although they were not given incentives to avoid smoking non-study commercial cigarettes, they were encouraged to honestly report their smoking behavior and were allowed to complete the trial regardless of compliance. During Week 6 of the study, smokers randomized to the lowest nicotine condition had significantly reduced tobacco use, dependence, and nicotine exposure compared to the usual brand and normal nicotine control conditions.

The results of CENIC-p1 provide empirical support for nicotine reduction as a regulatory strategy, but they must be interpreted cautiously due to substantial noncompliance to randomized treatment assignment. For example, among participants randomized to smoke very low nicotine content (VLNC) cigarettes (cigarettes with 0.4 mg of nicotine per gram of tobacco), 39% reported smoking at least 1 non-study cigarette during week 6, and 80% reported smoking at least 1 non-study cigarette at some point during the trial. A per protocol analysis, that is, analyzing only compliant participants, is problematic because compliance status is confounded, and compliers may differ systematically from non-compliers. The primary analysis of CENIC-p1 followed the intention-to-treat (ITT) principle and analyzed the data from all participants according to their randomized treatment assignment regardless of their compliance. An ITT analysis provides an unbiased estimator of the effect of a treatment or intervention when it is used in an environment (e.g., target population, level of non-compliance, etc.) similar to the clinical trial environment ([Hernán and Hernández-Díaz, 2012](#)). However, if the nicotine content of cigarettes were limited by regulation, and smokers no longer had legal access to standard commercial cigarettes, the effect of nicotine reduction on smoking behavior may be different than in the trial.

Our goal is to estimate the effect of smoking VLNC cigarettes on cigarette consumption and other measures of smoking behavior (dependence, withdrawal, etc.) in the hypothetical world where a regulatory body has reduced the nicotine content of cigarettes, and normal nicotine content commercial cigarettes are no longer available. In the language of clinical trials, we wish to estimate the effect of smoking VLNC cigarettes in the presence of complete compliance, that is, the causal effect ([Bellamy and others, 2007](#)). Methods for estimating the causal effect of an intervention in the presence of noncompliance are well-established and include inverse probability of compliance weighted (IPCW) estimators ([Hernán and Robins, 2006](#); [Cain and Cole, 2009](#)), principal stratification ([Frangakis and Rubin, 2002](#)), structural nested models estimated by G-estimation ([Robins, 1994](#)), and instrumental variable approaches ([Angrist and others, 1996](#)).

Existing methods for estimating causal effects in the presence of noncompliance assume that investigators know, without error, whether or not a participant was compliant. In the context of randomized tobacco trials, like CENIC-p1, self-reported compliance is subject to error and recall bias, and analyzing biomarkers of nicotine exposure has been suggested as an alternate approach to identify non-compliant participants (Benowitz and others, 2015). One recent study (Denlinger and others, 2016) evaluating biomarkers of nicotine exposure in participants exclusively smoking cigarettes with 0.4 mg per gram of tobacco found that the 95th percentile for total nicotine equivalents (TNEs, a biomarker of short-term nicotine exposure that measures most nicotine metabolites) was 6.41 nmol/mL. Yet among the CENIC-p1 participants randomized to the 0.4 mg per gram of tobacco arm, 63% of participants who self-reported full compliance during Week 6 had TNE greater than 6.41 nmol/mL (Nardone and others, 2016). This demonstrates that self-reported compliance can substantially misclassify whether or not participants were compliant to their randomized treatment assignment. Although certain biomarkers (e.g., TNE) may suggest that a participant’s self-reported compliance status is incorrect, no biomarker of nicotine exposure perfectly discriminates between compliers and non-compliers.

We propose a novel estimator of the causal effect from a randomized clinical trial when compliance status is measured with error. In contrast to existing methods, our estimator explicitly accounts for the potential for misclassification of compliers and non-compliers. Although we treat compliance status as an unobserved variable, we show how to weight participants by the product of their probability of compliance given the observed data and the inverse probability of compliance given confounders, resulting in a consistent and asymptotically normal estimator of the causal effect. Our simulation results illustrate that in finite samples our estimator outperforms *ad hoc* causal methods which ignore the error in compliance status by using either self-reported data or by using an estimated indicator of compliance. When there is a perfect discriminator of compliers and non-compliers, our estimator reduces to the standard IPCW estimator.

2. CAUSAL EFFECT ESTIMATORS

2.1. Potential outcomes and target of inference

We consider a hypothetical randomized clinical trial, where $A = 1, 2, \dots, r$ denotes the treatment group, which we assume is randomized. Let Z be a measure (without error) of noncompliance with $Z = 0$ indicating full compliance and increasing Z indicating greater noncompliance. For example, Z may be the number of non-study commercial cigarettes smoked in week 6 of CENIC-p1 or the number of pills not taken in a therapeutic clinical trial. Define the compliance indicator $C = I(Z = 0)$, and note that Z and C are not directly observed if noncompliance is measured with error. Define $Y^*(a, z)$ to be the outcome of a randomly-selected participant if, possibly contrary to fact, we set $A = a$ and $Z = z$. Because for each participant we do not observe $Y^*(a, z)$ for all a and z , $Y^*(a, z)$ is a potential outcome. For CENIC-p1, $Y^*(a, z)$ is the number of study cigarettes with nicotine content a smoked per day in week 6 of the study if the participant were to smoke z non-study cigarettes per day. The target of inference is $\mu(a, 0) - \mu(a', 0) = E\{Y^*(a, 0) - Y^*(a', 0)\}$, the expected difference of the outcome among randomized 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 (Pearl, 2001), that is, the treatment effect when the mediator, number of non-study cigarettes, is set at the fixed value of 0.

2.2. Observed data

Let Y be the observed outcome for a randomly-selected participant. Define the self-reported compliance indicator variable D , with $D = 1$ indicating the participant reports full compliance and $D = 0$ indicating that the participant reports any noncompliance. Let X be a vector of patient variables, and let B be a

biomarker indicating exposure to the treatment A . In the context of CENIC-p1, Y is the number of study cigarettes smoked per day during week 6 of the trial, B is a biomarker of nicotine exposure (e.g., TNE or cotinine) measured during week 6, $D = 1$ if the participant reports smoking 0 non-study cigarettes during week 6, and $D = 0$ if the participant reports smoking any non-study cigarettes during week 6. Note that participants self-report D , and this may be subject to error and recall biases.

2.3. Identifying assumptions

To relate the distribution of the observed data to the distribution of the potential outcome $Y^*(a, 0)$ in the case where C is observed, we make the following identifying assumptions (Robins and Hernán, 2008). First, we assume that we have measured enough covariates such that the compliance status is conditionally ignorable. That is, we assume the probability that a participant is compliant depends only on the observed covariates X and A and not additionally on any potential outcomes; this implies that $E\{C|A, X, Y^*(a, 0)\} = E(C|A, X)$ (no unmeasured confounders assumption). Although there may be additional variables associated with C , for example, the biomarker B , we assume that there are no additional confounders aside from X . Second, we assume that $E(C|A, X) > 0$ for all X and A , that is, there is positive probability of complying with the randomized treatment assignment within all levels of the confounders (positivity assumption). Finally, we assume that if $Z = 0$ and $A = a$, then $Y = Y^*(a, 0)$, that is, if a participant is compliant with the trial protocol, then her observed outcome is the same as it would be if the participant were forced to be compliant with the trial protocol (consistency assumption). When C is unobserved, additional assumptions will be necessary and will be stated where required.

2.4. Proposed estimators

We have stressed that C may not be directly observed due to measurement error or misclassification, but assume for now that $E(C|A, B, X, Y, D)$ and $E(C|A, X)$ are known. We discuss in Section 3 how these expectations can be estimated even though C is unobserved. We can estimate $\mu(a, 0)$ by solving the following estimating equation:

$$\sum_{i=1}^n \frac{E(C_i|A_i, B_i, X_i, Y_i, D_i)}{E(C_i|A_i, X_i)} \{Y_i - \mu(a, 0)\} I(A_i = a) = 0. \quad (2.1)$$

A similar estimator could be constructed for $\mu(a', 0)$, and the difference in those estimators could be used to estimate the treatment effect.

The estimating function with weights $\frac{E(C|A, B, X, Y, D)}{E(C|A, X)}$ is similar to a standard IPCW estimator with weights $\frac{C}{E(C|A, X)}$, but we have replaced the numerator with the conditional expected value of compliance instead of an indicator variable for compliance status. Note that if all participants self-reported compliance without error, or if there exists a biomarker that can perfectly discriminate compliers from non-compliers, then $E(C|A, B, X, Y, D)$ equals 0 or 1, and the estimating function in Equation (2.1) simplifies to the standard IPCW estimating function. Although it is not obvious that we can obtain valid inference without observing C , we show in Section 4 that, under suitable regularity conditions, $\hat{\mu}(a, 0)$, the solution to the estimating equation, is a consistent and asymptotically normal estimator of $\mu(a, 0)$. Because our estimator has an expectation, rather than an indicator variable, in the numerator to account for potential misclassification, we refer to the estimator as the Compliance Unsure RE-weighted estimator, or CURE estimator.

3. ESTIMATING THE WEIGHTS

In practice, $E(C|A, B, X, Y, D)$ and $E(C|A, X)$ are usually unknown and must be estimated. To estimate these, we begin by re-writing $E(C|A, B, X, Y, D)$ as a function of components that can be estimated directly from the observed data. We then show how auxiliary data, if available, can be used to obtain a more precise estimate of $E(C|A, B, X, Y, D)$. Finally, we discuss how $\hat{E}(C|A, B, X, Y, D)$, the estimate of $E(C|A, B, X, Y, D)$, can be used to estimate $E(C|A, X)$.

3.1. Estimating the numerator of the weights

First, using Bayes' Theorem, we can write $E(C|A = a, B = b, X = x, Y = y, D = d)$ as:

$$\frac{f(b|a, x, y, d, c = 1; \xi) \cdot \rho(a, x, y, d; \alpha)}{f(b|a, x, y, d, c = 1; \xi) \cdot \rho(a, x, y, d; \alpha) + f(b|a, x, y, d, c = 0; \xi) \cdot \{1 - \rho(a, x, y, d; \alpha)\}} \quad (3.1)$$

where f is the conditional density of B given A, X, Y, D , and C indexed by parameter vector ξ , and $\rho(a, x, y, d; \alpha) = \Pr(C = 1|A = a, X = x, Y = y, D = d; \alpha)$ indexed by parameter vector α . Rewriting the expectation shifts the goal from estimating the conditional expectation of C to estimating the conditional distribution of $B|A, X, Y, D, C$. Note that B, A, X, Y , and D are all observed random variables, which allows us to directly estimate the density of $B|A, X, Y, D$ as the mixture density

$$g(b|a, x, y, d; \xi, \alpha) = \rho(a, x, y, d; \alpha) \cdot f(b|a, x, y, d, c = 1; \xi) + \{1 - \rho(a, x, y, d; \alpha)\} \cdot f(b|a, x, y, d, c = 0; \xi). \quad (3.2)$$

Thus, although we do not observe C , we can estimate $E(C|A, B, X, Y, D)$ by estimating the conditional density of $B|A, X, Y, D$. The maximum likelihood estimators of ξ and α solve the score equations

$$\sum_{i=1}^n \frac{\partial}{\partial (\xi^T, \alpha^T)^T} \log g(B_i|A_i, X_i, Y_i, D_i; \xi, \alpha) = 0. \quad (3.3)$$

As is frequently the case with mixture distributions, Equation (3.3) may be difficult to solve directly. In that case, we can find the maximum likelihood estimates using the expectation-maximization (EM) algorithm (Dempster and others, 1977). In Section A of the supplementary material available at *Biostatistics* online, we give details of the EM algorithm updates for ξ and α .

Equations (3.1)–(3.3) can sometimes be simplified based on the scientific problem. For example, in some applications it may be reasonable to simplify the modeling assumptions of the conditional density of B . In particular, according the directed acyclic graph (DAG) in Section B of the supplementary materials available at *Biostatistics* online, which is one plausible DAG for the CENIC-p1 data, B may be conditionally independent of X given A, Y , and C . We note that this assumption is testable using observed data so the data analyst does not need to a priori assume the correct causal structure. Additionally, it may be reasonable to assume that if $D = 0$, that is, the participant reports noncompliance with the study protocol, then $C = 0$ without error. That is, participants may not erroneously report noncompliance because in most trials there is usually no incentive to be non-compliant. This would imply that $E(C|A, B, X, Y, D = 0) = 0$.

3.2. Incorporating compliance information from an auxiliary study

Estimating mixture distributions is challenging in practice due to difficulties in identifying the underlying component densities. An advantage to estimating the mixture distribution of $B|A, X, Y, D$ for CENIC-p1

is the presence of an auxiliary study to evaluate biomarkers of nicotine exposure in fully compliant participants. [Denlinger and others \(2016\)](#) present data on biomarkers of nicotine exposure for 23 smokers who volunteered to be sequestered in a hotel for four nights with access to only cigarettes with 0.4 mg nicotine per gram of tobacco. These participants are known to be compliant, and their data can be used in conjunction with the data from CENIC-p1 to estimate the density of the biomarker in compliers. This will enhance our ability to identify the underlying components of the mixture distribution for $B|A, X, Y, D$. While these auxiliary data may seem unique to CENIC-p1, similar data arise in other settings. For example, pharmacokinetic/pharmacodynamic data from early-phase clinical trials could be used to help identify compliers in a phase III therapeutic clinical trial. In our application, we do not want to include these individuals' outcomes, Y , in estimation of the causal effect due to differences in smoking behavior between participants in the auxiliary and primary studies, but we do expect the distribution of $B|A, X, Y, D = 1, C = 1$ to be consistent across studies. This is known as the *transportability* assumption in the measurement error literature ([Carroll and others, 2006](#)).

Let n_k denote the number of participants in the auxiliary study, let $m = n + n_k$, and define the indicator variable K that equals 1 if the participant is included in the auxiliary data or 0 if the participant is included in the main trial. We can incorporate the auxiliary data when estimating ξ and α by solving the score equations

$$\sum_{i=1}^m \frac{\partial}{\partial (\xi^T, \alpha^T)^T} \{ (1 - K_i) \cdot \log g(B_i|A_i, X_i, Y_i, D_i; \xi, \alpha) + K_i \cdot \log f(B_i|A_i, X_i, Y_i, d = 1, c = 1; \xi) \} = 0. \quad (3.4)$$

3.3. Estimating the denominator of the weights

In most applications, the denominator of the weights will be unknown and must be estimated. Because the denominator must be between 0 and 1, we specify a regression model $E(C_i|A_i, X_i) = \pi_i(\beta) = g^{-1} \{ \beta_0 + X_i^T \beta_1 + I(A_i = 2)\beta_2 + \dots + I(A_i = r)\beta_r \}$, 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. If $E(C|A, B, X, Y, D)$ were known, we could estimate β by solving the estimating equation

$$\sum_{i=1}^m (1 - K_i) \frac{E(C_i|A_i, B_i, X_i, Y_i, D_i; \xi, \alpha) - \pi_i(\beta)}{\pi_i(\beta) \{1 - \pi_i(\beta)\}} \frac{\partial \pi_i(\beta)}{\partial \beta^T} = 0. \quad (3.5)$$

That is, $E(C|A, B, X, Y, D; \xi, \alpha)$ is the ‘‘response’’ of the regression model. Note that β can be estimated using standard software, for example, using the `glm` function in R. Because $E(C|A, B, X, Y, D; \xi, \alpha)$ is between 0 and 1, this is analogous to modeling proportions in a logistic or probit model. In the case that ξ and α are unknown, we can stack Equations (3.4) and (3.5) and solve jointly, which is equivalent to replacing $E(C|A, B, X, Y, D; \xi, \alpha)$ in Equation (3.5) with the estimated expectation $E(C|A, B, X, D, Y, D; \hat{\xi}, \hat{\alpha})$.

4. ASYMPTOTIC PROPERTIES OF THE CURE ESTIMATOR

In discussing the asymptotic properties of the proposed estimator, for simplicity we assume there are no auxiliary data of the type described in Section 3.2 and consider only a single-arm trial with $a = 1$ for all participants, but the results easily generalize to multi-arm trials. Under the assumptions in Section 2.3, the estimating function has expectation equal to 0. The key to demonstrating this is to note that in expectation, our proposed estimator is equivalent to an IPCW estimator in which the compliance status is known without error:

$$\begin{aligned}
E \left[\frac{E(C_i|B_i, X_i, Y_i, D_i)}{E(C_i|X_i)} \{Y_i - \mu(1, 0)\} \right] &= E \left[\frac{C_i}{E(C_i|X_i)} \{Y_i - \mu(1, 0)\} \right] \\
&= E \left[\frac{E\{C_i|X_i, Y_i^*(1, 0)\}}{E(C_i|X_i)} \{Y_i^*(1, 0) - \mu(1, 0)\} \right] \\
&= E \left[\frac{E(C_i|X_i)}{E(C_i|X_i)} \{Y_i^*(1, 0) - \mu(1, 0)\} \right] = 0. \tag{4.1}
\end{aligned}$$

The 1st equality follows from iterated expectation. At this step, the argument of the expectation is now equivalent to the case when C is known without error, and the remaining equalities follow from assumptions stated in Section 2.3: the 2nd follows from the consistency assumption and iterated expectation, and the 3rd follows from the no unmeasured confounders assumption. Note that the result in Equation (4.1) hold even if B is null. That is, we do not need to measure a biomarker of exposure and could replace $E(C_i|B_i, X_i, Y_i, D_i)$ with $E(C_i|X_i, Y_i, D_i)$ and the estimating function would still have expectation equal to zero. Nevertheless, conditioning on B allows us to incorporate auxiliary compliance data as described in Section 3.2. Furthermore, including a biomarker of compliance improves the discrimination between the compliers and non-compliers which will improve the computational stability of the estimates in the mixture distribution in Equation (3.2).

We can simultaneously estimate $\mu(1, 0)$, ξ , α , and β by stacking the estimating equations given in Equations (2.1), (3.4), and (3.5). Let $\mathbf{Z}_i = (A_i, B_i, X_i, Y_i, D_i)$ be the observed data on participant i and $\sum_{i=1}^n \psi_i \{\mathbf{Z}_i; \mu(1, 0), \xi, \alpha, \beta\}$ denote the stacked estimating function. We showed above that the first component of $\psi_i \{\mathbf{Z}_i; \mu(1, 0), \xi, \alpha, \beta\}$ has expectation 0; the components corresponding to $(\xi^T, \alpha^T)^T$ and β have expectation 0 due to being score functions of a log likelihood and of a generalized linear model, respectively. The fact that the stacked estimating function has expectation 0 implies that, under suitable regularity conditions,

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

where $V = E([\psi_i \{\mathbf{Z}_i; \mu(1, 0), \xi, \alpha, \beta\}] [\psi_i \{\mathbf{Z}_i; \mu(1, 0), \xi, \alpha, \beta\}]^T)$ and $U = -E \left[\frac{\partial \psi_i \{\mathbf{Z}_i; \mu(1, 0), \xi, \alpha, \beta\}}{\partial \{\mu(1, 0), \xi^T, \alpha^T, \beta^T\}^T} \right]$. The sandwich covariance matrix $U^{-1}V(U^{-1})^T$ can be estimated using the empirical averages for U and V (see, e.g., [Stefanski and Boos \(2002\)](#)) or with the bootstrap ([Efron, 1979](#)).

Note that unlike IPCW estimators when C is known where only the model for $E(C_i|X_i)$ must be correctly specified, the CURE estimator also relies on correctly specifying the distribution of the mixture components in Equation (3.2) to obtain consistent and asymptotically normal estimators.

We compare the relative efficiency of the CURE estimator to the standard IPCW estimator when compliance is observed in order to understand the consequences of measuring compliance with error. If $E(C|X)$ and $E(C|B, X, Y, D)$ are known and do not need to be estimated, then the CURE estimator is more efficient than the standard IPCW estimator with weights $\frac{C}{E(C|X)}$. To see this, first note that $E \left(\frac{\partial}{\partial \mu(1, 0)} \left[\frac{C_i}{E(C_i|X_i)} \{Y_i - \mu(1, 0)\} \right] \right) = E \left(\frac{\partial}{\partial \mu(1, 0)} \left[\frac{E(C_i|B_i, X_i, Y_i, D_i)}{E(C_i|X_i)} \{Y_i - \mu(1, 0)\} \right] \right)$, so the difference in the limiting variance of the estimators is due to differences in the variances of the estimating function. Next, we can write

$$\text{var} \left[\frac{C_i}{E(C_i|X_i)} \{Y_i - \mu(1, 0)\} \right] = \text{var} \left[\frac{C_i - E(C_i|B_i, X_i, Y_i, D_i) + E(C_i|B_i, X_i, Y_i, D_i)}{E(C_i|X_i)} \{Y_i - \mu(1, 0)\} \right].$$

Then, because

$$\begin{aligned} & \text{cov} \left[\frac{C_i - E(C_i|B_i, X_i, Y_i, D_i)}{E(C_i|X_i)} \{Y_i - \mu(1, 0)\}, \frac{E(C_i|B_i, X_i, Y_i, D_i)}{E(C_i|X_i)} \{Y_i - \mu(1, 0)\} \right] \\ &= E \left(\left[\frac{E(C_i|B_i, X_i, Y_i, D_i) - E(C_i|B_i, X_i, Y_i, D_i)}{E(C_i|X_i)} \{Y_i - \mu(1, 0)\} \right] \frac{E(C_i|B_i, X_i, Y_i, D_i)}{E(C_i|X_i)} \{Y_i - \mu(1, 0)\} \right) \\ &= 0 \end{aligned}$$

by iterated expectation, the variance of the estimating function with weights $\frac{C}{E(C|X)}$ is

$$\begin{aligned} & \text{var} \left[\frac{C_i}{E(C_i|X_i)} \{Y_i - \mu(1, 0)\} \right] \\ &= \text{var} \left[\frac{E(C_i|B_i, X_i, Y_i, D_i)}{E(C_i|X_i)} \{Y_i - \mu(1, 0)\} \right] + \text{var} \left[\frac{C_i - E(C_i|B_i, X_i, Y_i, D_i)}{E(C_i|X_i)} \{Y_i - \mu(1, 0)\} \right] \\ &\geq \text{var} \left[\frac{E(C_i|B_i, X_i, Y_i, D_i)}{E(C_i|X_i)} \{Y_i - \mu(1, 0)\} \right]. \end{aligned}$$

Thus, if all expectations are known and do not need to be estimated, the CURE estimator is more efficient asymptotically than the standard IPCW estimator. To gain some intuition for why this occurs, consider the case where compliance has no effect on the outcome: the CURE estimator gains efficiency by simply taking a sample average of the outcome among participants randomized to group a to estimate $\mu(a, 0)$, whereas the IPCW estimator excludes participants who are non-compliant even though compliance has no effect on the outcome. As the effect of compliance on the outcome strengthens, $E(C|B, X, Y, D)$ approaches an indicator function and the CURE estimator approaches the standard IPCW estimator. In general, the CURE estimator borrows more data from the non-compliers as the effect of compliance decreases, which increases efficiency over the IPCW estimator.

A natural question is how estimating the weights impacts the variances of the estimators. Here, we use notation consistent with the generalized linear model notation introduced previously, and we remind the reader that $E(C|X)$ can be written as $\pi(\beta)$. When the weights are $\frac{C}{\pi(\beta)}$, and β must be estimated jointly with $\mu(1, 0)$, the asymptotic variance of $\hat{\mu}(1, 0)$ is

$$E \left[\frac{\{Y_i^*(1, 0) - \mu(1, 0)\}^2}{\pi_i(\beta)} \right] - H_1 H_2^{-1} H_1^T,$$

where $H_1 = E \left[\frac{Y_i^*(1, 0) - \mu(1, 0)}{\pi_i(\beta)} \frac{\partial \pi_i(\beta)}{\partial \beta} \right]$ and $H_2 = E \left[\frac{1}{\pi_i(\beta)(1 - \pi_i(\beta))} \frac{\partial \pi_i(\beta)}{\partial \beta^T} \frac{\partial \pi_i(\beta)}{\partial \beta} \right]$. This interesting result shows that the variance of $\hat{\mu}(1, 0)$ with weights $\frac{C}{\pi(\beta)}$ is reduced when β is estimated compared to when β is known (Lunceford and Davidian, 2004). When the weights are $\frac{E(C|B, X, Y, D; \xi, \alpha)}{\pi(\beta)}$, when ξ, α , and β must be estimated jointly with $\mu(1, 0)$, and when there are no auxiliary data to use in estimating the parameters of the mixture density (3.2), then the large sample variance of $\hat{\mu}(1, 0)$ is

$$\begin{aligned} & E \left[\frac{E(C_i|B_i, X_i, Y_i, D_i)^2}{\pi_i(\beta)^2} \{Y_i - \mu(1, 0)\}^2 \right] + 2 \left[(-H_3 H_4^{-1} + H_1 H_2^{-1} H_5 H_4^{-1}) \quad (-H_1 H_2^{-1}) \right] \begin{bmatrix} H_6^T \\ H_7^T \end{bmatrix} \\ &+ \left[(-H_3 H_4^{-1} + H_1 H_2^{-1} H_5 H_4^{-1}) \quad (-H_1 H_2^{-1}) \right] \begin{bmatrix} H_4 & H_8 \\ H_8^T & H_9 \end{bmatrix} \begin{bmatrix} (-H_3 H_4^{-1} + H_1 H_2^{-1} H_5 H_4^{-1})^T \\ (-H_1 H_2^{-1})^T \end{bmatrix}, \quad (4.2) \end{aligned}$$

where

$$\begin{aligned}
H_3 &= -E \left[\frac{Y_i - \mu(1, 0)}{\pi_i(\beta)} \frac{\partial E(C_i|B_i, X_i, Y_i, D_i)}{\partial(\xi^T, \alpha^T)} \right] \\
H_4 &= E \left[\frac{\partial \log g(B_i|X_i, Y_i, D_i; \xi, \alpha)}{\partial(\xi^T, \alpha^T)^T} \frac{\partial \log g(B_i|X_i, Y_i, D_i; \xi, \alpha)}{\partial(\xi^T, \alpha^T)} \right] \\
H_5 &= -E \left[\pi_i(\beta) \{1 - \pi_i(\beta)\} \frac{\partial \pi_i(\beta)}{\partial \beta^T} \frac{\partial E(C_i|B_i, X_i, Y_i, D_i)}{\partial(\xi^T, \alpha^T)} \right] \\
H_6 &= E \left[\frac{C_i \{Y_i^*(1, 0) - \mu(1, 0)\}}{\pi_i(\beta)} \frac{\partial \log g(B_i|X_i, Y_i, D_i; \xi, \alpha)}{\partial(\xi^T, \alpha^T)} \right] \\
H_7 &= E \left[\frac{E(C_i|B_i, X_i, Y_i, D_i) \{Y_i - \mu(1, 0)\} \{E(C_i|B_i, X_i, Y_i, D_i) - \pi_i(\beta)\}}{\pi_i(\beta)^2 \{1 - \pi_i(\beta)\}} \frac{\partial \pi_i(\beta)}{\partial \beta} \right] \\
H_8 &= E \left[\frac{E(C_i|B_i, X_i, Y_i, D_i) - \pi_i(\beta)}{\pi_i(\beta) \{1 - \pi_i(\beta)\}} \frac{\partial \log g(B_i|X_i, Y_i, D_i; \xi, \alpha)}{\partial(\xi^T, \alpha^T)^T} \frac{\partial \pi_i(\beta)}{\beta} \right] \\
H_9 &= E \left[\frac{\{E(C_i|B_i, X_i, D_i, Y_i) - \pi_i(\beta)\}^2}{[\pi_i(\beta) \{1 - \pi_i(\beta)\}]^2} \frac{\partial \pi_i(\beta)}{\partial \beta^T} \frac{\partial \pi_i(\beta)}{\partial \beta} \right]
\end{aligned}$$

and H_1 and H_2 were defined previously. The last term in (4.2) is positive because it is a quadratic form of a positive semi-definite matrix, but the middle term does not have this property and is neither clearly positive nor clearly negative under all conditions. Thus, although estimating the weights for the IPCW estimator is guaranteed to increase the asymptotic efficiency of $\hat{\mu}(1, 0)$, there is no such guarantee for the CURE estimator. If C were observed and $E(C|B, X, Y, D)$ were estimated using a regression model in the CURE estimator, then we could make a definitive statement about the impact of estimating $(\xi^T, \alpha^T)^T$ on the limiting variance of $\mu(1, 0)$, but this is not possible when C is not observed and we used the approach described in Section 3. Also note that the information matrix H_4 will contain more information when auxiliary data are incorporated as described in Section 3.2, but this may be beneficial only for estimation of the mixture density (3.2) in small samples, not for asymptotic efficiency of $\hat{\mu}(1, 0)$.

5. SIMULATION STUDY

We designed a simulation study to test the finite-sample properties of the proposed estimator. For simplicity, we consider a scenario in which all participants are assigned a single treatment, $A = 1$, and only estimate $\mu(1, 0)$, rather than a difference in means between two treatment groups. To facilitate data generation, define C' as a latent continuous measure of compliance. We generated (C', X, Y) from a multivariate normal

distribution with mean vector $(0, 10, 16)^T$ and covariance matrix $\Sigma = \begin{pmatrix} 1.000 & -0.573 & -0.139 \\ -0.573 & 2.000 & 1.715 \\ -0.139 & 1.715 & 3.000 \end{pmatrix}$.

We define the compliance indicator $C = I \{C' > \Phi^{-1}(0.80)\}$, where Φ is the standard normal cumulative distribution function. This gives $Pr(C = 1) = 0.20$, consistent with the preliminary estimates of the compliance rate from CENIC-p1 (Nardone and others, 2016). We simulated $B \sim N(-9.3 - 0.8 \cdot C + 0.7 \cdot Y, \sigma^2)$ with two values of σ^2 , 0.818 and 0.663, to give an area under the receiver operating characteristic (ROC) curve of 0.8 or 0.9, respectively, for discriminating between compliers and non-compliers. We let $D = H^{1-C}$, where H is a Bernoulli random variable with success probability 0.3 independent of other data, so that those truly compliant ($C = 1$) always self-report compliance and for non-compliant participants ($C = 0$) H is an indicator of whether or not non-compliance was reported with error. The data generated

are consistent with the DAG shown in the supplementary material Section B available at *Biostatistics* online, which is one possible DAG for the CENIC-p1 data. Note that the DAG implies that B and X are conditionally independent given C and Y (and A , which here has only one level).

We considered sample sizes of 225 (roughly the number of participants randomized to the VLNC cigarettes in CENIC-p1), 500, and 1,000 and also included data from an auxiliary study in which participants are known to be compliant as described in Section 3.2 equal to 10% of the size of the main clinical trial.

We compared five estimators of $\mu(1, 0)$: (i) a per protocol estimator based on self-reported compliance; (ii) an inverse probability weighted (IPW) estimator based on self-reported compliance with weights equal to $\frac{D}{\hat{E}(D|X)}$; (iii) an IPW estimator with weights equal to $\frac{C}{\hat{E}(C|X)}$; (iv) a cutoff IPW estimator that first estimates $E(C|B, X, Y, D)$ using the methods described in Section 3.1, defines $\hat{C} = \hat{E}(C|B, X, Y, D) > 0.5$, and then uses the weights $\frac{\hat{C}}{\hat{E}(\hat{C}|X)}$; and (v) the CURE estimator with weights $\frac{\hat{E}(C|B, X, Y, D)}{\hat{E}(C|X)}$. Although we have argued the IPW estimator cannot be used in CENIC-p1, we include it for comparison to illustrate the cost of relying on imperfect measures of compliance. We assumed that $E(C|X, Y, D = 0) = 0$, and observations with $D = 0$ did not contribute to estimation of the mixture distribution. To estimate $E(C|B, X, Y, D = 1)$ in the cutoff IPW and CURE estimators, we assumed simple linear regression models with normal residuals for $B|Y, D = 1, C = 1$ and $B|Y, D = 1, C = 0$; to estimate $E(C|X, Y, D = 1)$ we assumed a generalized linear model with probit link. To improve computation time, we used a single set of starting values for the EM algorithm which were estimated parameters from models fit using the actual compliance C . In applications, when C is unknown to the analyst and cannot be used to generate starting values, the EM algorithm may require multiple iterations with different starting values to find the global maximum of the likelihood, but the simulation nevertheless gave good results using only this one set of starting values for each iteration. For the self-reported IPW, cutoff IPW, and CURE estimators, the denominator of the weights was estimated using a generalized linear model with probit link where the outcome is the numerator of the weights. We used the bootstrap percentile method with 1,000 bootstrap re-sampled data sets to compute 95% confidence intervals (CI).

Table 1 shows the simulation results. Overall, the CURE estimator has very small bias for both area under the curve (AUC) of 0.8 and 0.9. With AUC of 0.8, the CURE estimator has higher mean squared error than per protocol or IPW based on self-report, but with sample size 1,000 the mean squared error is smaller; with AUC of 0.9, the CURE estimator has lower mean squared error for sample sizes 500 and 1,000. The per protocol and self-report IPW estimators show bias that is not attenuated with increasing sample size and coverage probabilities which are not close to the nominal level. The cutoff IPW estimator has low bias, but, surprisingly, for each sample size and AUC level, the CURE estimator has much smaller mean squared error and coverage probability closer to the nominal 0.95 level. Unsurprisingly, the CURE estimator has higher mean squared error than the IPW estimator, but the mean squared error of the CURE estimator approaches that of the IPW estimator as the AUC increases. The simulation results demonstrate that the CURE estimator has better small sample performance than per protocol and self-report IPW when there is potential for misclassification. Furthermore, the CURE estimator performs better than an *ad hoc* estimator that uses IPW with an estimated indicator variable of compliance.

6. APPLICATION TO THE CENIC-P1 DATA

We applied the CURE estimator to estimate the causal effect of VLNC cigarettes on the number of cigarettes smoked per day using data from CENIC-p1. Although CENIC-p1 was a 6-week trial, for simplicity we are only concerned with compliance and outcomes collected in the last week. In this analysis, we let $A = 1$ if the participant was randomized to smoke VLNC cigarettes (0.4 mg nicotine per gram of tobacco, high

Table 1. *Simulation results*

n	Estimator	Bias	MC SD	Mean SE	CP	MSE
225	Per Protocol	-0.629	0.173	0.174	0.045	0.426
	Self-Report IPW	-0.403	0.155	0.154	0.257	0.187
	IPW	-0.035	0.330	0.278	0.905	0.110
	Cutoff IPW _{AUC=0.8}	-0.140	0.860	0.740	0.978	0.758
	CURE _{AUC=0.8}	-0.119	0.768	0.689	0.976	0.603
	Cutoff IPW _{AUC=0.9}	-0.037	0.605	0.574	0.972	0.367
	CURE _{AUC=0.9}	-0.033	0.523	0.518	0.970	0.274
500	Per Protocol	-0.636	0.114	0.117	0.000	0.417
	Self-Report IPW	-0.405	0.101	0.103	0.024	0.174
	IPW	-0.015	0.209	0.196	0.926	0.044
	Cutoff IPW _{AUC=0.8}	-0.010	0.665	0.659	0.983	0.442
	CURE _{AUC=0.8}	-0.047	0.530	0.571	0.978	0.283
	Cutoff IPW _{AUC=0.9}	0.021	0.368	0.419	0.982	0.136
	CURE _{AUC=0.9}	-0.021	0.277	0.344	0.965	0.077
1000	Per Protocol	-0.633	0.086	0.082	0.000	0.408
	Self-Report IPW	-0.405	0.071	0.072	0.000	0.169
	IPW	-0.009	0.154	0.143	0.937	0.024
	Cutoff IPW _{AUC=0.8}	0.070	0.480	0.531	0.983	0.235
	CURE _{AUC=0.8}	-0.011	0.325	0.418	0.979	0.106
	Cutoff IPW _{AUC=0.9}	0.053	0.244	0.279	0.977	0.062
	CURE _{AUC=0.9}	-0.007	0.173	0.208	0.963	0.030

Subscripts indicate the area under the ROC curve for discriminating compliers from non-compliers. MC SD, Monte Carlo standard deviation of the estimator; Mean SE, mean estimated standard error of the estimator; CP, coverage probability of 95% CI, MSE: mean squared error.

and low tar groups combined) and let $A = 2$ if randomized to smoke usual brand cigarettes. All other notation in this application is defined in Section 2.

The goal of this analysis is to estimate the causal contrast $\mu(2, 0) - \mu(1, 0)$, the expected reduction in cigarettes smoked per day during week 6 if smoking only VLNC cigarettes versus smoking usual brand cigarettes. We estimate the causal effect by estimating $\mu(2, 0)$ and $\mu(1, 0)$ separately and taking their difference. The usual brand group is meant to represent smoking behavior with commercially available cigarettes and, in that sense, participants in this group were never treated as non-compliant, and $\mu(2, 0)$ was estimated using the sample average of the total number of cigarettes smoked per day during week 6 (i.e., study plus non-study cigarettes). We consider the four estimators for $\mu(1, 0)$ discussed in the simulation study (excluding IPW because C is unobserved) and include the ITT estimator for comparison.

As in the simulation, we assumed that $E(C|A, X, Y, D = 0) = 0$, and participants with $D = 0$ did not contribute to estimation of the mixture distribution. Using the biomarker $\log(\text{TNE})$ measured at week 6 as the (only) biomarker B of exposure, we estimated the probability of compliance for participants self-reporting compliance (i.e., $D = 1$) following the approach in Section 3.1. Specifically, in fitting the mixture distribution in Equation (3.2), we assume a simple linear regression model with normally distributed errors for $B|A = 1, Y, D = 1, C$ with no shared parameters between the different levels of C . As in the simulation, we assume that B and X are conditionally independent given A, Y , and C , consistent with the DAG in Section B of the supplementary material available at *Biostatistics* online. We assumed a logistic regression model for $\Pr(C = 1|A = 1, X, Y, D = 1)$, where the confounders X included age, level

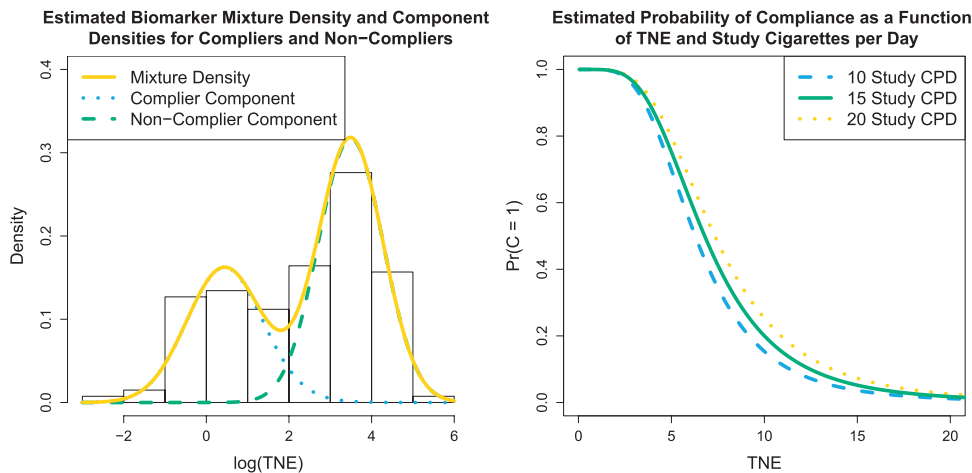


Fig. 1. Left: histogram of $\log(\text{TNE})$, the biomarker of nicotine exposure, for the self-reported compliers randomized to smoke cigarettes with 0.4 mg nicotine per gram tobacco in the CENIC-p1 trial. The estimated mixture distribution (solid line) and the distribution of each of the mixture components (dashed and dotted lines) is overlaid. Right: probability a participant who self-reported compliance was actually compliant as a function of TNE and number of self-reported study cigarettes smoked per day.

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). We incorporated the data for the 23 participants from [Denlinger and others \(2016\)](#) who were known to be compliant to aid in estimating the parameters of the mixture distribution as described in Section 3.2. Multiple sets of starting values were tried for the EM algorithm, and we used those values which gave the lowest negative log likelihood. We estimated the denominator $\Pr(C = 1|A = 1, X)$ following the approach of Section 3.3 using a logit link with the same predictors X described above. All CIs were estimated using the non-parametric bootstrap percentile method with 1,000 bootstrap resamples.

For the VLNC group, 137 of 222 (61.7%) participants self-reported compliance during week 6. The left panel of Figure 1 shows a histogram of $B = \log(\text{TNE})$ for the self-reported compliers in the treatment group, the estimated mixture distribution, and the complier and non-complier component distributions, which supports our parametric assumptions for the components of the mixture distribution. The right panel shows the probability of compliance as a function of TNE and Y , the self-reported number of cigarettes per day. We estimated $\Pr(C = 1|A = 1, D = 1) = 0.376$, and $\Pr(C = 1|A = 1) = 0.224$, indicating a substantial proportion of self-reported compliers were non-compliant. Estimated coefficients and parameters for the numerator and denominator of the weights and 95% bootstrap CIs can be found in Section C of the supplementary material available at *Biostatistics* online. We also include a table giving some summary statistics of baseline characteristics, confounders, and the biomarker Week 6 $\log(\text{TNE})$.

Table 2 shows the estimated causal effect of VLNC cigarettes on number of cigarettes smoked per day. The cutoff IPW gives the most optimistic estimate of the causal effect of the treatment, while the CURE estimate is more conservative. In contrast, the per protocol and self-report IPW estimators give similar and more modest estimates of the treatment effect. Although the CURE and cut-off IPW estimates are similar, note that the length of the 95% CI is much wider for the cut-off IPW estimator than for the CURE estimator.

Table 2. Point estimates, standard error (SE) of the estimators, and 95% CI 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)
Causal estimators					
Per protocol	22.18	15.12	7.07	1.80	(3.61, 10.63)
Self-report IPW	22.18	15.19	6.99	1.67	(3.66, 10.38)
Cutoff IPW	22.18	14.83	7.35	3.03	(1.64, 12.99)
CURE	22.18	14.98	7.20	2.79	(2.01, 12.46)

$\mu(2, 0)$: 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.

We typically expect the ITT estimator to be more conservative than estimators of the causal effect. Although the ITT estimator was in fact more conservative than the causal estimator, the difference is small considering the large proportion of noncompliance. While this may seem counterintuitive, the impact of non-compliance may be different compared to other clinical trials of medication. In the case of medication, we expect a monotone dose-response relationship, and non-compliance with the medication should dilute the treatment effect by reducing the dose received. Here, on the one hand, non-compliant use of high nicotine cigarettes could actually reduce the need for study cigarettes more than the study cigarette itself (e.g., by more effectively alleviating withdrawal). Consequently, one might expect the number of study cigarettes smoked per day to be lower in non-compliant participants than it would be if they were forced to be compliant. On the other hand, non-compliance is also associated with individuals who find VLNC cigarettes particularly unsatisfying. Such individuals might be less inclined to continue to smoke or would smoke less if forced to be compliant.

The results presented here require us to assume that we have correctly modeled the numerator and denominator of the weights and should be interpreted cautiously. Like all models, the assumptions must be considered when interpreting outcomes, and convergent analyses should be used to clarify the likely mechanism whenever possible. Finally, it is important to note that in a regulatory environment in which VLNC cigarettes were the only legally available cigarettes, we would expect that the proportion of smokers using only VLNC cigarettes would be substantially higher than in CENIC-p1, but there would still likely be some use of cigarettes with higher nicotine content (e.g., hoarding, black market).

7. DISCUSSION

Methods for estimating causal effects from randomized clinical trials when there is noncompliance frequently rely on imperfect measures of compliance. Estimators that do not acknowledge the error in the measures of compliance will result in biased estimators of the causal effect. We developed a causal estimator that accounts for uncertainty in compliance status by re-weighting a typical IPCW estimator by a participant's probability of compliance given a biomarker of compliance, the outcome of interest, and confounders. Although we treated the true compliance status as unobserved, we showed the probability of compliance can be estimated by assuming the distribution of the biomarker follows a mixture distribution with separate components for compliers and non-compliers. The simulation demonstrates that our proposed estimator has little bias, good coverage probability, and smaller mean squared error than an *ad hoc* estimator.

The methods developed here have particular relevance to and were motivated by regulatory tobacco research. There is usually substantial noncompliance in regulatory tobacco trials due to the availability of commercial tobacco products. The causal analysis using the methods we have developed, as compared to an ITT analysis, is likely to better estimate the effect we would observe if regulations changing the nicotine composition in cigarettes were enacted. However, our proposed method also has broad applicability for clinical trials conducted in other therapeutic areas. The method is particularly attractive in cases where investigators rely on imperfect measures of compliance, such as participants' self-report (e.g., pill counts, timeline follow back, etc.) because the method explicitly accounts for the uncertainty of compliance status.

The preceding has assumed that either the outcome Y is not subject to measurement/self-report error or that one is interested in the average causal effect on self-reported outcomes. We show in Section D of the supplementary material available at *Biostatistics* online that if the observed outcome is subject to measurement error, the proposed approach will estimate the causal effect if there were no measurement error under mild assumptions.

Others have investigated the effect of and possible solutions to mediation estimators when the mediator (e.g., compliance to randomized treatment group) is measured with error (Valeri and others, 2014; Ogburn and VanderWeele, 2012). Most prior work has examined the effect on regression-based estimators as opposed to our IPW framework. Additionally, our approach makes minimal assumptions about measurement error. In particular, we do not need to assume that self-reported compliance (i.e., the covariate measured with error) is a surrogate for true compliance for valid inference. That is, the method does not require that D is conditionally independent of Y given C or that D is conditionally independent of Y given C and X .

There are several limitations to our approach. First, estimating the parameters of the mixture distributions may be computationally challenging, resulting in unstable parameter estimates. In our simulation, we relied on an auxiliary data set that included data from participants whose compliance was known. However, such data sources are frequently available in other settings as well, such as in pharmacokinetic/pharmacodynamic studies. Second, we only considered compliance and outcomes during week 6 of the CENIC-p1 trial. This was done mainly for simplicity, however, and we could develop a longitudinal extension of the estimator, which is a likely subject of future work. Finally, inverse probably weighted estimators are known to be inefficient. The efficiency of the CURE estimator could likely be improved through an augmented weighted estimator (Tsiatis, 2006).

Causal inference methods frequently rely on poor measures of compliance. Our causal estimator weights participants by the product of their probability of compliance given the biomarker of treatment exposure and the inverse probability of compliance given confounders. Our approach suggests that, rather than improving methods of eliciting compliance status from participants, perhaps a more fruitful of area of research is in developing biomarkers of exposure. We restricted our attention to IPCW-like weights and developed the method for a point exposure study, but future work may develop causal estimators in other settings. Our hope is that the proposed methods becomes a standard analysis by investigators estimating causal effects from clinical trials.

8. SOFTWARE

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

SUPPLEMENTARY MATERIAL

Supplementary material is available at <http://biostatistics.oxfordjournals.org>.

ACKNOWLEDGMENTS

Conflict of Interest: None declared.

FUNDING

This research was partially funded by NIH grants 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

- ANGRIST, J., IMBENS, G. W. AND RUBIN, D. B. (1996). Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association* **91**, 444–472.
- BELLAMY, S. L., LIN, J. Y. AND TEN H., THOMAS R. (2007). An introduction to causal modeling in clinical trials. *Clinical Trials* **4**, 58–73.
- BENOWITZ, N. L. AND HENNINGFIELD, J. E. (1994). Establishing a nicotine threshold for addiction: the implications for tobacco regulation. *The New England Journal of Medicine* **331**, 123–125.
- BENOWITZ, N. L., NARDONE, N., HATSUKAMI, D. K. AND DONNY, E. C. (2015). Biochemical estimation of noncompliance with smoking of very low nicotine content cigarettes. *Cancer Epidemiology, Biomarkers & Prevention* **24**, 331–335.
- CAIN, L. E. AND COLE, S. R. (2009). Inverse probability-of-censoring weights for the correction of time-varying noncompliance in the effect of randomized highly active antiretroviral therapy on incident aids or death. *Statistics in Medicine* **28**, 1725–1738.
- CARROLL, R. J., RUPPERT, D., STEFANSKI, L. A. AND CRAINICEANU, C. M. (2006). *Measurement Error in Nonlinear Models*, 2nd edition. Boac Raton, FL: Chapman & Hall/CRC.
- DEMPSTER, A. P., LAIRD, N. M. AND RUBIN, D. B. (1977). Maximum likelihood from incomplete data via EM algorithm. *Journal of the Royal Statistical Society Series B-Methodological* **39**, 1–38.
- DENLINGER, R. L., SMITH, T. T., MURPHY, S. E., KOOPMEINERS, J. S., BENOWITZ, N. L., HATSUKAMI, D. K., PACEK, L. R., COLINO, C., CWALINA, S. N. AND DONNY, E. C. (2016). Nicotine and anatabine exposure from very low nicotine content cigarettes. *Tobacco Regulatory Science* **2**, 186–203.
- DONNY, E. C., DENLINGER, R. L., TIDEY, J. W., KOOPMEINERS, J. S., BENOWITZ, N. L., VANDREY, R. G., AL'ABSI, M., CARMELLA, S. G., CINCIRIPINI, P. M., DERMODY, S. S., and others. (2015). Randomized trial of reduced-nicotine standards for cigarettes. *The New England Journal of Medicine* **373**, 1340–9.
- EFRON, B. (1979). Bootstrap methods: Another look at the Jackknife. *The Annals of Statistics* **7**, 1–26.
- FRANGAKIS, C. E. AND RUBIN, D. B. (2002). Principal stratification in causal inference. *Biometrics* **58**, 21–29.
- HERNÁN, M. A. AND HERNÁNDEZ-DÍAZ, S. (2012). Beyond the intention-to-treat in comparative effectiveness research. *Clinical Trials* **9**, 48–55.
- HERNÁN, M. A. AND ROBINS, J. M. (2006). Estimating causal effects from epidemiological data. *Journal of Epidemiology and Community Health* **60**, 578–586.
- 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.
- NARDONE, N., DONNY, E. C., HATSUKAMI, D. K., KOOPMEINERS, J. S., MURPHY, S. E., STRASSER, A. A., TIDEY, J. W., VANDREY, R. AND BENOWITZ, N. L. (2016). Estimations and predictors of non-compliance in switchers to reduced nicotine content cigarettes. *Addiction* Doi: 10.1111/add.13519.

- OGBURN, E. L. AND VANDERWEELE, T. J. (2012). Analytic results on the bias due to nondifferential misclassification of a binary mediator. *American Journal of Epidemiology* **176**, 555–561.
- PEARL, J. (2001). Direct and indirect effects. In: *Proceedings of the Seventeenth Conference on Uncertainty and Artificial Intelligence* (TECHNICAL REPORT, R-273-UAI). San Francisco: Morgan Kaufmann, pp. 411–420.
- ROBINS, J. M. (1994). Correcting for noncompliance in randomized trials using structural nested mean models. *Communications in Statistics-Theory and Methods* **23**, 2379–2412.
- ROBINS, J. M. AND HERNÁN, M. A. (2008). *Longitudinal Data Analysis*, Chapter 23: Estimation of the Causal Effects of Time-Varying Exposures. Boca Raton: Chapman & Hall/CRC.
- STEFANSKI, L. A. AND BOOS, D. D. (2002). The calculus of m-estimation. *The American Statistician*, 29–38.
- TSIATIS, A. A. (2006). *Semiparametric Theory and Missing Data*. New York: Springer Science+Business Media, LLC.
- U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES. (2014). *The Health Consequences of Smoking - 50 Years of Progress: A Report of the Surgeon General*. Atlanta, GA: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health.
- VALERI, L., LIN, X. AND VANDERWEELE, T. J. (2014). Mediation analysis when a continuous mediator is measured with error and the outcome follows a generalized linear model. *Statistics in Medicine* **33**, 4875–4890.

[Received November 2, 2016; revised March 21, 2017; accepted for publication May 7, 2017]