ipwErrorY: An R Package for Estimation of Average Treatment Effect with Misclassified Binary Outcome

by Di Shu and Grace Y. Yi

Abstract It has been well documented that ignoring measurement error may result in severely biased inference results. In recent years, there has been limited but increasing research on causal inference with measurement error. In the presence of misclassified binary outcome variable, Shu and Yi (2017) considered the inverse probability weighted estimation of the average treatment effect and proposed valid estimation methods to correct for misclassification effects for various settings. To expedite the application of those methods for situations where misclassification in the binary outcome variable is a real concern, we implement correction methods proposed by Shu and Yi (2017) and develop an R package **ipwErrorY** for general users. Simulated datasets are used to illustrate the use of the developed package.

Introduction

Causal inference methods have been widely used in empirical research (e.g., Rothman et al., 2008; Imbens and Rubin, 2015; Hernán and Robins, 2019). The propensity score, defined to be the probability of an individual to receive the treatment, plays an important role in conducting causal inference (Rosenbaum and Rubin, 1983). Many causal inference methods have been developed based on the propensity score (e.g., Rosenbaum, 1987, 1998; Robins et al., 2000; Lunceford and Davidian, 2004). These methods commonly require modeling the treatment assignment, which can be difficult in some applications. To protect against misspecification of the treatment model, various methods have been proposed (e.g., Robins et al., 1994; Scharfstein et al., 1999; Bang and Robins, 2005). Among them, doubly robust methods are often advocated since the resulting estimators are still consistent when either the treatment model or the outcome model (but not both) is misspecified; such an attractive property is referred to as double robustness.

Although many methods are available for causal inference such as for the estimation of average treatment effects (ATE), those methods are vulnerable to poor quality data. Typically when data are error-contaminated, most existing methods would be inapplicable. It has been well documented that measurement error in variables can often lead to seriously biased inference results in many settings (e.g., Fuller, 1987; Gustafson, 2003; Carroll et al., 2006; Buonaccorsi, 2010; Yi, 2017).

In the context of causal inference with error-prone data, there has been limited but increasing research on the impact of measurement error on causal inference and the development of correction methods to deal with measurement error. For instances, McCaffrey et al. (2013) proposed a correction estimation method when baseline covariates are error-prone. Babanezhad et al. (2010) examined the bias arising from ignoring misclassification in the treatment variable. Braun et al. (2016) developed a correction method to correct for treatment misclassification using validation data.

In settings with misclassification in the binary outcome variable, Shu and Yi (2017) explored the estimation of ATE using the inverse probability weighted (IPW) method. They derived the asymptotic bias caused by misclassification and developed consistent estimation methods to eliminate the misclassification effects. Their development covers practical scenarios where (1) the misclassification probabilities are known, or (2) the misclassification probabilities are unknown but validation data or replicates of outcome measurements are available for their estimation. They further propose a doubly robust estimator to provide protection against possible misspecification of the treatment model.

The methods developed by Shu and Yi (2017) enjoy wide applications, because misclassified binary outcome data arise commonly in practice. For example, the self-reported smoking status without being confirmed by biochemical tests is subject to misclassification; results of screening tests are often subject to false positive error and/or false negative error. For datasets with outcome misclassification, ignoring misclassification effects may lead to severely biased results. To expedite the application of the correction methods for general users, we develop an R package, called **ipwErrorY** (Shu and Yi, 2019), to implement the methods by Shu and Yi (2017) for practical settings where the commonly-used logistic regression model is employed for the treatment model and the outcome model. The package focuses on measurement error in the *outcome Y* only but not on other types of measurement error, such as measurement error in covariates.

The remainder is organized as follows. First, we introduce the notation and framework. Secondly,

we describe the methods to be implemented in R. Thirdly, we present the implementation steps and illustrate the use of the package with examples. Finally, a discussion is given. The developed R package **ipwErrorY** is available from the Comprehensive R Archive Network (CRAN) at https://CRAN.R-project.org/package=ipwErrorY.

Notation and framework

Let Y_1 be the binary potential outcome that would have been observed had the individual been treated, and Y_0 be the binary potential outcome that would have been observed had the individual been untreated. Let X be the vector of baseline covariates; T be the observed binary treatment variable; and Y be the observed binary outcome.

Being consistent with the usual causal inference framework (e.g., Lunceford and Davidian, 2004), we make the following standard assumptions:

Assumption 1 (Consistency): $Y = TY_1 + (1 - T)Y_0$;

Assumption 2 (No Unmeasured Confounding): $(Y_1, Y_0) \perp\!\!\!\perp T|X$;

Assumption 3 (Positivity): $0 < P(T = 1|X) < 1$.

The objective is to estimate the ATE, defined as $\tau_0 = E(Y_1) - E(Y_0)$. Suppose we have a sample of size n. For $i = 1, \cdots, n$, we add subscript i to notations X, T and Y to denote the corresponding variable for individual i.

In the presence of outcome misclassification, instead of Y, its surrogate version Y^* is observed. We assume that

$$P(Y^* = a|Y = b, X, T) = P(Y^* = a|Y = b) \quad \text{for} \quad a, b = 0, 1. \tag{1}$$

That is, conditional on the true value of the outcome, the misclassification probability is assumed to be homogeneous for all the individuals, regardless of their covariate information or treatment status. For ease of exposition, write $p_{ab} = P(Y^* = a|Y = b)$. Then the sensitivity and the specificity are expressed by p_{11} and p_{00}, respectively, and the misclassification probabilities are $p_{01} = 1 - p_{11}$ and $p_{10} = 1 - p_{00}$.

Estimation methods

In this section we present estimation methods for τ_0 under three scenarios. We first start with the case where misclassification probabilities are given, and then consider settings where misclassification probabilities are unknown but can be estimated by using additional data sources.

Estimation with known misclassification probabilities

In this subsection we assume that misclassification probabilities p_{11} and p_{10} are known. For $i = 1, \ldots, n$, let $e_i = P(T_i = 1|X_i)$ be the conditional probability for individual i to receive the treatment, also termed as the propensity score (e.g., Rosenbaum and Rubin, 1983), a quantity that plays an important role in causal inference. To correct for outcome misclassification effects, Shu and Yi (2017) proposed an estimator of τ_0 given by

$$\widehat{\tau} = \frac{1}{p_{11} - p_{10}} \left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{T_i Y_i^*}{\widehat{e}_i} - \frac{1}{n} \sum_{i=1}^{n} \frac{(1 - T_i)Y_i^*}{1 - \widehat{e}_i} \right\}, \tag{2}$$

where $\widehat{e}_i$ is an estimate of the propensity score $e_i = P(T_i = 1|X_i)$ obtained by fitting the treatment model relating T to X.

The estimator $\widehat{\tau}$, given by (2), is a consistent estimator of τ_0, provided regularity conditions including Assumptions 1-3. The sandwich variance estimate of $\widehat{\tau}$ can be obtained using the theory of estimating functions (e.g., Newey and McFadden, 1994; Heyde, 1997; Yi, 2017, Ch.1).

To estimate the propensity score e_i for $i = 1, \ldots, n$, we specifically characterize e_i using the widely-used logistic regression model. That is, the treatment model is given by

$$\text{logit } P(T_i = 1|X_i) = \gamma_0 + \gamma_X^\top X_i, \tag{3}$$

for $i = 1, \ldots, n$, where $\gamma = (\gamma_0, \gamma_X^\top)^\top$ is the vector of parameters. As a result, an unbiased estimating

function of γ is taken as the score function

$$\left\{ T_i - \frac{1}{1 + \exp(-\gamma_0 - \gamma_X^\top X_i)} \right\} (1, X_i^\top)^\top. \tag{4}$$

Let $\theta = (\tau, \gamma^\top)^\top$. Shu and Yi (2017) showed that

$$\Psi(Y_i^*, T_i, X_i; \theta) = \begin{pmatrix} \left\{ T_i - \dfrac{1}{1 + \exp(-\gamma_0 - \gamma_X^\top X_i)} \right\} (1, X_i^\top)^\top \\[2ex] \dfrac{T_i Y_i^*}{e_i} - \dfrac{(1 - T_i)Y_i^*}{1 - e_i} - (p_{11} - p_{10})\tau \end{pmatrix} \tag{5}$$

is an unbiased estimating function of θ. Solving $\sum_{i=1}^{n} \Psi(Y_i^*, T_i, X_i; \theta) = 0$ for θ yields an estimator of θ, denoted by $\widehat{\theta}$.

Let $\theta_0 = (\tau_0, \gamma_0^\top)^\top$ be the true value of θ. Define $A(\theta_0) = E\left\{ -(\partial/\partial\theta^\top)\Psi(Y^*, T, X; \theta)|_{\theta=\theta_0} \right\}$ and $B(\theta_0) = E\{\Psi(Y^*, T, X; \theta)\Psi^\top(Y^*, T, X; \theta)|_{\theta=\theta_0}\}$. Under regularity conditions, we have that

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N\left(0, A(\theta_0)^{-1}B(\theta_0)A(\theta_0)^{-1\top} \right) \quad \text{as } n \to \infty. \tag{6}$$

Consequently, the variance of $\widehat{\theta}$ can be estimated by the empirical sandwich estimator:

$$\widehat{Var}(\widehat{\theta}) = \frac{1}{n} A_n(\widehat{\theta})^{-1} B_n(\widehat{\theta}) A_n(\widehat{\theta})^{-1\top}, \tag{7}$$

where

$$A_n(\widehat{\theta}) = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta^\top} \Psi(Y_i^*, T_i, X_i; \theta)|_{\theta=\widehat{\theta}} \tag{8}$$

and

$$B_n(\widehat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \Psi(Y_i^*, T_i, X_i; \theta)\Psi^\top(Y_i^*, T_i, X_i; \theta)|_{\theta=\widehat{\theta}}. \tag{9}$$

Then the variance estimate of $\widehat{\tau}$ in (2) is given by $\widehat{Var}(\widehat{\tau}) = \widehat{v}_{11}$, where $\widehat{v}_{11}$ is the element of the first row and the first column of $\widehat{Var}(\widehat{\theta})$. A $(1 - \alpha)100\%$ confidence interval of τ_0 is given by $\widehat{\tau} \pm z_{\alpha/2} \times \sqrt{\widehat{Var}(\widehat{\tau})}$, where α is a specified value between 0 and 1, and $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution.

Estimation with validation data

In this subsection we assume that misclassification probabilities p_{11} and p_{10} are unknown and that there is an internal validation subsample $\mathcal{V}$ of size n_v which collects measurements of variables X, T, Y and Y^*.

With the validation data, p_{11} and p_{10} can be estimated as

$$\widehat{p}_{11} = \frac{\sum_{i \in \mathcal{V}} Y_i Y_i^*}{\sum_{i \in \mathcal{V}} Y_i} \quad \text{and} \quad \widehat{p}_{10} = \frac{\sum_{i \in \mathcal{V}}(1 - Y_i)Y_i^*}{\sum_{i \in \mathcal{V}}(1 - Y_i)}, \tag{10}$$

respectively.

To estimate τ_0, one may use error-free outcome data of the validation subsample to construct an estimator

$$\widehat{\tau}_\text{v} = \frac{1}{n_\text{v}} \sum_{i \in \mathcal{V}} \frac{T_i Y_i}{\widehat{e}_i} - \frac{1}{n_\text{v}} \sum_{i \in \mathcal{V}} \frac{(1 - T_i)Y_i}{1 - \widehat{e}_i}, \tag{11}$$

of τ_0, or alternatively, one may apply (2) to estimate τ_0 using non-validation data with the resulting estimator given by

$$\widehat{\tau}_\text{N} = \frac{1}{\widehat{p}_{11} - \widehat{p}_{10}} \left\{ \frac{1}{n - n_\text{v}} \sum_{i \notin \mathcal{V}} \frac{T_i Y_i^*}{\widehat{e}_i} - \frac{1}{n - n_\text{v}} \sum_{i \notin \mathcal{V}} \frac{(1 - T_i)Y_i^*}{1 - \widehat{e}_i} \right\}, \tag{12}$$

where $\widehat{p}_{11}$ and $\widehat{p}_{10}$ are given by (10).

Although the validation data based estimator $\widehat{\tau}_\text{v}$ given by (11) and the non-validation data based estimator $\widehat{\tau}_\text{N}$ given by (12) are both consistent estimators of τ_0, they both incur efficiency loss due to

the inability of utilizing all the available data.

Shu and Yi (2017) considered the linear combination of $\widehat{\tau}_V$ and $\widehat{\tau}_N$

$$\widehat{\tau}(c) = c\widehat{\tau}_V + (1-c)\widehat{\tau}_N, \tag{13}$$

where c is a constant between 0 and 1.

For any c, the consistency of $\widehat{\tau}(c)$ is immediate due to the consistency of $\widehat{\tau}_V$ and $\widehat{\tau}_N$. However, the efficiency of $\widehat{\tau}(c)$ depends on the choice of c. Typically, $Var\{\widehat{\tau}(c)\}$ is minimized at

$$c_{OPT} = \frac{Var(\widehat{\tau}_N) - Cov(\widehat{\tau}_V, \widehat{\tau}_N)}{Var(\widehat{\tau}_V) + Var(\widehat{\tau}_N) - 2Cov(\widehat{\tau}_V, \widehat{\tau}_N)}, \tag{14}$$

suggesting that $\widehat{\tau}(c_{OPT}) = c_{OPT}\widehat{\tau}_V + (1 - c_{OPT})\widehat{\tau}_N$ is the optimal estimator among the linear combination estimators formulated as (13). Furthermore, c_{OPT} can be estimated by

$$\widehat{c}_{OPT} = \frac{\widehat{Var}(\widehat{\tau}_N) - \widehat{Cov}(\widehat{\tau}_V, \widehat{\tau}_N)}{\widehat{Var}(\widehat{\tau}_V) + \widehat{Var}(\widehat{\tau}_N) - 2\widehat{Cov}(\widehat{\tau}_V, \widehat{\tau}_N)}, \tag{15}$$

where $\widehat{Var}(\widehat{\tau}_N)$, $\widehat{Cov}(\widehat{\tau}_V, \widehat{\tau}_N)$ and $\widehat{Var}(\widehat{\tau}_V)$ are the estimates for $Var(\widehat{\tau}_N)$, $Cov(\widehat{\tau}_V, \widehat{\tau}_N)$ and $Var(\widehat{\tau}_V)$, respectively.

To obtain $\widehat{Var}(\widehat{\tau}_V)$, $\widehat{Cov}(\widehat{\tau}_V, \widehat{\tau}_N)$ and $\widehat{Var}(\widehat{\tau}_N)$, Shu and Yi (2017) constructed an unbiased estimating function by combining the estimating functions (11) and (12), where they introduced different symbols, say τ_V and τ_N, to denote the parameter τ for which (11) and (12), respectively, are used to estimate; both τ_V and τ_N have the true value τ_0. Let $\theta = (\tau_V, \tau_N, \gamma^\top, p_{11}, p_{10})^\top$. Define

$$\Psi_c(Y_i^*, T_i, X_i, Y_i; \theta) = \begin{pmatrix} \left\{ T_i - \dfrac{1}{1 + \exp(-\gamma_0 - \gamma_X^\top X_i)} \right\} (1, X_i^\top)^\top \\[2mm] (Y_i Y_i^* - p_{11} Y_i) \cdot I(i \in \mathcal{V}) \cdot \dfrac{n}{n_V} \\[2mm] \{(1 - Y_i)Y_i^* - p_{10}(1 - Y_i)\} \cdot I(i \in \mathcal{V}) \cdot \dfrac{n}{n_V} \\[2mm] \left\{ \dfrac{T_i Y_i}{e_i} - \dfrac{(1 - T_i)Y_i}{1 - e_i} - \tau_V \right\} \cdot I(i \in \mathcal{V}) \cdot \dfrac{n}{n_V} \\[2mm] \left\{ \dfrac{T_i Y_i^*}{e_i} - \dfrac{(1 - T_i)Y_i^*}{1 - e_i} - (p_{11} - p_{10})\tau_N \right\} I(i \notin \mathcal{V}) \cdot \dfrac{n}{n - n_V} \end{pmatrix}, \tag{16}$$

where $I(\cdot)$ is the indicator function. Then $\Psi_c(Y_i^*, T_i, X_i, Y_i; \theta)$ is an unbiased combined estimating function of θ. Solving $\sum_{i=1}^n \Psi_c(Y_i^*, T_i, X_i, Y_i; \theta) = 0$ for θ yields an estimator of θ, denoted by $\widehat{\theta}$. The variance of $\widehat{\theta}$ can be estimated by the empirical sandwich estimator, denoted as $\widehat{Var}(\widehat{\theta})$. Let $\widehat{v}_{i,j}$ be the element of the ith row and the jth column of $\widehat{Var}(\widehat{\theta})$. Then $\widehat{Var}(\widehat{\tau}_V) = \widehat{v}_{1,1}$, $\widehat{Cov}(\widehat{\tau}_V, \widehat{\tau}_N) = \widehat{v}_{1,2}$, and $\widehat{Var}(\widehat{\tau}_N) = \widehat{v}_{2,2}$.

Finally, Shu and Yi (2017) pointed out the two associated conditions: $Var(\widehat{\tau}_V) + Var(\widehat{\tau}_N) - 2Cov(\widehat{\tau}_V, \widehat{\tau}_N) \geq 0$ and $0 \leq c \leq 1$. If one or both conditions are violated with empirical estimates, $\widehat{c}_{OPT}$ is then set to be 1 if $\widehat{\tau}_V$ has smaller variance than $\widehat{\tau}_N$ and 0 otherwise. The resulting optimal linear combination estimator $\widehat{\tau}(\widehat{c}_{OPT})$ is

$$\widehat{\tau}_{OPT} = \widehat{c}_{OPT}\widehat{\tau}_V + (1 - \widehat{c}_{OPT})\widehat{\tau}_N, \tag{17}$$

with the variance estimate given by

$$\widehat{Var}(\widehat{\tau}_{OPT}) = \{\widehat{Var}(\widehat{\tau}_V) + \widehat{Var}(\widehat{\tau}_N) - 2\widehat{Cov}(\widehat{\tau}_V, \widehat{\tau}_N)\}\widehat{c}_{OPT}^2 - \{2\widehat{Var}(\widehat{\tau}_N) - 2\widehat{Cov}(\widehat{\tau}_V, \widehat{\tau}_N)\}\widehat{c}_{OPT} + \widehat{Var}(\widehat{\tau}_N). \tag{18}$$

A $(1 - \alpha)100\%$ confidence interval of τ_0 is given by $\widehat{\tau}_{OPT} \pm z_{\alpha/2} \times \sqrt{\widehat{Var}(\widehat{\tau}_{OPT})}$, where α is a specified value between 0 and 1, and $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution.

Estimation with replicates

In this subsection we assume that misclassification probabilities p_{11} and p_{10} are unknown and that two repeated outcome measurements are available for each individual. Suppose $Y_i^*(1)$ and $Y_i^*(2)$ are two independent replicates of Y_i. Let η denote the prevalence $P(Y = 1)$ and π_r be the probability of

obtaining r outcome observations equal to 1 among two repeated outcome measurements for $r = 0, 1$. Then

$$\pi_0 = \eta(1 - p_{11})^2 + (1 - \eta)(1 - p_{10})^2; \tag{19}$$

$$\pi_1 = 2\eta(1 - p_{11})p_{11} + 2(1 - \eta)(1 - p_{10})p_{10}. \tag{20}$$

Let $\theta = (\tau, \gamma^\top, \eta, p_{11}, p_{10})^\top$. Shu and Yi (2017) considered an unbiased estimating function of θ, given by

$$\Psi_r\{Y_i^*(1), Y_i^*(2), T_i, X_i; \theta\} = \begin{pmatrix} \left\{ T_i - \dfrac{1}{1 + \exp(-\gamma_0 - \gamma_X^\top X_i)} \right\} (1, X_i^\top)^\top \\[2ex] \{1 - Y_i^*(1)\} \cdot \{1 - Y_i^*(2)\} - \pi_0 \\[2ex] Y_i^*(1) \cdot \{1 - Y_i^*(2)\} + Y_i^*(2) \cdot \{1 - Y_i^*(1)\} - \pi_1 \\[2ex] \dfrac{T_i Y_i^*}{e_i} - \dfrac{(1 - T_i)Y_i^*}{1 - e_i} - (p_{11} - p_{10})\tau \end{pmatrix}, \tag{21}$$

where $Y_i^* = \{Y_i^*(1) + Y_i^*(2)\}/2$, together with a constraint imposed for achieving parameters identifiability (e.g., White et al., 2001; Yi and He, 2017).

Let $\widehat{\tau}_R$ denote the estimator of τ_0 obtained by solving

$$\sum_{i=1}^n \Psi_r\{Y_i^*(1), Y_i^*(2), T_i, X_i; \theta\} = 0 \tag{22}$$

for θ. The variance of $\widehat{\tau}_R$ can be estimated by the empirical sandwich estimator $\widehat{Var}(\widehat{\tau}_R)$. A $(1 - \alpha)100\%$ confidence interval of τ_0 is given by $\widehat{\tau}_R \pm z_{\alpha/2} \times \sqrt{\widehat{Var}(\widehat{\tau}_R)}$, where α is a specified value between 0 and 1, and $z_{\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal distribution.

Finally, we comment that when implementing (22), one of the following constraints is often used in applications: (C1) sensitivity equals specificity (i.e., $p_{11} = p_{00}$), (C2) sensitivity p_{11} is known, (C3) specificity p_{00} is known, and (C4) prevalence η is known. These four constraints are implemented in our R package.

Choosing a suitable identifiability constraint is primarily driven by the nature of data. When the false positive rate p_{10} and the false negative rate p_{01} are close, it is reasonable to impose the constraint that the sensitivity equals the specificity. When there is prior information on the value of the sensitivity, the specificity, or the prevalence, it is plausible to add the identifiability constraint (C2), (C3) or (C4). For example, in smoking cessation studies, patients who quit smoking (with $Y = 1$) are unlikely to report that they still smoke, so it may be reasonable to set the constraint $p_{11} = 1$. Sometimes, researchers may use the disease prevalence reported from another similar study for their own study, when such a prevalence is perceived to be close to that of the target population.

Doubly robust estimation

To protect against model misspecification, Shu and Yi (2017) proposed a doubly robust estimator of τ_0:

$$\widehat{\tau}_{DR} = \widehat{E}(Y_1) - \widehat{E}(Y_0), \tag{23}$$

where

$$\widehat{E}(Y_1) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{T_i Y_i^*}{\widehat{e}_i(p_{11} - p_{10})} - \frac{T_i - \widehat{e}_i}{\widehat{e}_i}\widehat{q}_{i1} - \frac{T_i}{\widehat{e}_i}\left(\frac{p_{10}}{p_{11} - p_{10}}\right) \right\}, \tag{24}$$

$$\widehat{E}(Y_0) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{(1 - T_i)Y_i^*}{(1 - \widehat{e}_i)(p_{11} - p_{10})} + \frac{T_i - \widehat{e}_i}{1 - \widehat{e}_i}\widehat{q}_{i0} - \frac{1 - T_i}{1 - \widehat{e}_i}\left(\frac{p_{10}}{p_{11} - p_{10}}\right) \right\}, \tag{25}$$

$\widehat{q}_{i1}$ is an estimate of $q_{i1} = P(Y_i = 1 | T_i = 1, X_i)$ and $\widehat{q}_{i0}$ is an estimate of $q_{i0} = P(Y_i = 1 | T_i = 0, X_i)$.

The estimator $\widehat{\tau}_{DR}$ enjoys the double robustness property in the sense that it is still consistent if one of the treatment model and the outcome model is incorrectly specified. In our developed R package, we particularly implement the following two scenarios.

Scenario 1 (Shared covariate effects for the treated and untreated groups):

Suppose the outcome model is postulated as

$$\text{logit } P(Y_i = 1|T_i, X_i) = \beta_0 + \beta_\text{T} T_i + \beta^\top X_i, \tag{26}$$

where β_0, β_T and β are the parameters. The model reflects the setting where the treated and untreated groups share the same covariate effect β on the outcome.

By (1) and (26), the observed likelihood function contributed from individual i is

$$L_i(\beta_0, \beta_\text{T}, \beta)$$

$$= P(Y_i^*|X_i, T_i)$$

$$= P(Y_i = 1|X_i, T_i)P(Y_i^*|X_i, T_i, Y_i = 1) + P(Y_i = 0|X_i, T_i)P(Y_i^*|X_i, T_i, Y_i = 0)$$

$$= \frac{1}{1 + \exp\{-\beta_0 - \beta_\text{T} T_i - \beta^\top X_i\}} \cdot \{p_{11} Y_i^* + (1 - p_{11})(1 - Y_i^*)\}$$

$$+ \frac{\exp\{-\beta_0 - \beta_\text{T} T_i - \beta^\top X_i\}}{1 + \exp\{-\beta_0 - \beta_\text{T} T_i - \beta^\top X_i\}} \cdot \{p_{10} Y_i^* + (1 - p_{10})(1 - Y_i^*)\}. \tag{27}$$

With regularity conditions, maximizing the observed likelihood $\prod_{i=1}^n L_i(\beta_0, \beta_\text{T}, \beta)$ with respect to $(\beta_0, \beta_\text{T}, \beta^\top)^\top$ gives a consistent estimator of $(\beta_0, \beta_\text{T}, \beta^\top)^\top$, denoted as $(\widehat{\beta}_0, \widehat{\beta}_\text{T}, \widehat{\beta}^\top)^\top$. It follows that q_{i1} and q_{i0}, are, respectively, estimated by

$$\widehat{q}_{i1} = \frac{1}{1 + \exp(-\widehat{\beta}_0 - \widehat{\beta}_\text{T} - \widehat{\beta}^\top X_i)} \tag{28}$$

and

$$\widehat{q}_{i0} = \frac{1}{1 + \exp(-\widehat{\beta}_0 - \widehat{\beta}^\top X_i)}. \tag{29}$$

Scenario 2 (Possibly different covariate effects for the treated and untreated groups):

Suppose that the outcome model is postulated as

$$\text{logit } P(Y_i = 1|T_i = 1, X_i) = \beta_{01} + \beta_1^\top X_i \tag{30}$$

for the treated group and

$$\text{logit } P(Y_i = 1|T_i = 0, X_i) = \beta_{00} + \beta_0^\top X_i \tag{31}$$

for the untreated group, where the parameters $(\beta_{01}, \beta_1^\top)^\top$ for the treated group may differ from the parameters $(\beta_{00}, \beta_0^\top)^\top$ for the untreated group.

To obtain a consistent estimator $(\widehat{\beta}_{01}, \widehat{\beta}_1^\top)^\top$ of $(\beta_{01}, \beta_1^\top)^\top$ and a consistent estimator $(\widehat{\beta}_{00}, \widehat{\beta}_0^\top)^\top$ of $(\beta_{00}, \beta_0^\top)^\top$, we employ the observed likelihood for the treated group and the untreated group separately. For example, the observed likelihood function contributed from individual l in the treated group (i.e., $T_l = 1$) is

$$L_{1,l}(\beta_{01}, \beta_1)$$

$$= P(Y_l^*|T_l = 1, X_l)$$

$$= P(Y_l = 1|T_l = 1, X_l)P(Y_l^*|Y_l = 1) + P(Y_l = 0|T_l = 1, X_l)P(Y_l^*|Y_l = 0)$$

$$= \frac{1}{1 + \exp\{-\beta_{01} - \beta_1^\top X_l\}} \cdot \{p_{11} Y_l^* + (1 - p_{11})(1 - Y_l^*)\}$$

$$+ \frac{\exp\{-\beta_{01} - \beta_1^\top X_l\}}{1 + \exp\{-\beta_{01} - \beta_1^\top X_l\}} \cdot \{p_{10} Y_l^* + (1 - p_{10})(1 - Y_l^*)\}. \tag{32}$$

Maximizing the observed likelihood $\prod_{l:T_l=1} L_{1,l}(\beta_{01}, \beta_1)$ with respect to β_{01} and β_1 gives us a consistent estimator $(\widehat{\beta}_{01}, \widehat{\beta}_1^\top)^\top$, provided regularity conditions. Similarly, we calculate the observed likelihood function $L_{0,k}(\beta_{00}, \beta_0)$ for individual k in the untreated group (i.e., $T_k = 0$), and then obtain the estimator $(\widehat{\beta}_{00}, \widehat{\beta}_0^\top)^\top$ by maximizing the observed likelihood $\prod_{l:T_k=0} L_{0,k}(\beta_{00}, \beta_0)$ with respect to

β_{00} and β_0. Thus, q_{i1} and q_{i0} are estimated by

$$\widehat{q}_{i1} = \frac{1}{1 + \exp(-\widehat{\beta}_{01} - \widehat{\beta}_1^\top X_i)} \tag{33}$$

and

$$\widehat{q}_{i0} = \frac{1}{1 + \exp(-\widehat{\beta}_{00} - \widehat{\beta}_0^\top X_i)}, \tag{34}$$

respectively.

Variance estimator of $\widehat{\tau}_{\mathrm{DR}}$:

Consistency and asymptotic normality of $\widehat{\tau}_{\mathrm{DR}}$ can be established using the theory of estimating functions. Below we derive the sandwich variance estimator of $\widehat{\tau}_{\mathrm{DR}}$ by constructing an unbiased estimating function using the "delta method" in the M-estimator framework (Stefanski and Boos, 2002).

Define β_{F} to be the vector of parameters for the outcome models. Under Scenario 1 with shared covariate effects for the treated and untreated groups, $\beta_{\mathrm{F}} = (\beta_0, \beta_{\mathrm{T}}, \beta^\top)^\top$. Under Scenario 2 with possibly different covariate effects for the treated and untreated groups, $\beta_{\mathrm{F}} = (\beta_{\mathrm{F1}}^\top, \beta_{\mathrm{F0}}^\top)^\top$ with $\beta_{\mathrm{F1}} = (\beta_{01}, \beta_1^\top)^\top$ and $\beta_{\mathrm{F0}} = (\beta_{00}, \beta_0^\top)^\top$. Let $\theta = (\gamma^\top, \beta_{\mathrm{F}}^\top, \mu_1, \mu_0, \tau)^\top$, where μ_1 and μ_0 represent $E(Y_1)$ and $E(Y_0)$, respectively.

We construct the following unbiased estimating function for θ:

$$\Psi_{dr}\{Y_i^*, T_i, X_i; \theta\} = \begin{pmatrix} \left\{ T_i - \dfrac{1}{1 + \exp(-\gamma_0 - \gamma_X^\top X_i)} \right\} (1, X_i^\top)^\top \\[2ex] \psi(Y_i^*, T_i, X_i; \beta_{\mathrm{F}}) \\[2ex] \left\{ \dfrac{T_i Y_i^*}{\widehat{e}_i(p_{11} - p_{10})} - \dfrac{T_i - \widehat{e}_i}{\widehat{e}_i}\widehat{q}_{i1} - \dfrac{T_i}{\widehat{e}_i}\left(\dfrac{p_{10}}{p_{11} - p_{10}}\right) \right\} - \mu_1 \\[2ex] \left\{ \dfrac{(1 - T_i) Y_i^*}{(1 - \widehat{e}_i)(p_{11} - p_{10})} + \dfrac{T_i - \widehat{e}_i}{1 - \widehat{e}_i}\widehat{q}_{i0} - \dfrac{1 - T_i}{1 - \widehat{e}_i}\left(\dfrac{p_{10}}{p_{11} - p_{10}}\right) \right\} - \mu_0 \\[2ex] \mu_1 - \mu_0 - \tau \end{pmatrix}, \tag{35}$$

where $\psi(Y_i^*, T_i, X_i; \beta_{\mathrm{F}})$ is the unbiased estimating equation for β_{F} derived from the observed likelihood. Specifically, under Scenario 1,

$$\psi(Y_i^*, T_i, X_i; \beta_{\mathrm{F}}) = \partial \log\{L_i(\beta_0, \beta_{\mathrm{T}}, \beta)\}/\partial(\beta_0, \beta_{\mathrm{T}}, \beta), \tag{36}$$

and under Scenario 2,

$$\psi(Y_i^*, T_i, X_i; \beta_{\mathrm{F}}) = (\psi_1(Y_i^*, T_i, X_i; \beta_{\mathrm{F1}})^\top, \psi_0(Y_i^*, T_i, X_i; \beta_{\mathrm{F0}})^\top)^\top \tag{37}$$

with

$$\psi_1(Y_i^*, T_i, X_i; \beta_{\mathrm{F1}}) = \partial \log\{L_{1,i}(\beta_{01}, \beta_1)\}/\partial(\beta_{01}, \beta_1) \cdot I(T_i = 1) \cdot \frac{n}{n_{\mathrm{T}}} \tag{38}$$

and

$$\psi_0(Y_i^*, T_i, X_i; \beta_{\mathrm{F0}}) = \partial \log\{L_{0,i}(\beta_{00}, \beta_0)\}/\partial(\beta_{00}, \beta_0) \cdot I(T_i = 0) \cdot \frac{n}{n - n_{\mathrm{T}}}, \tag{39}$$

where n_{T} is the size of the treated group.

By the theory of estimating functions (e.g., Newey and McFadden, 1994; Heyde, 1997; Yi, 2017, Ch.1), solving $\sum_{i=1}^n \Psi_{dr}\{Y_i^*, T_i, X_i; \theta\} = 0$ for θ yields an estimator of θ, denoted by $\widehat{\theta}$. Let θ_0 be the true value of θ. Define

$$A(\theta_0) = E\left\{ -(\partial/\partial\theta^\top)\Psi_{dr}(Y^*, T, X; \theta)|_{\theta=\theta_0} \right\}$$

and

$$B(\theta_0) = E\{\Psi_{dr}(Y^*, T, X; \theta)\Psi_{dr}^\top(Y^*, T, X; \theta)|_{\theta=\theta_0}\}.$$

Under regularity conditions, we have that

$$\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} N\left(0, A(\theta_0)^{-1}B(\theta_0)A(\theta_0)^{-1\top}\right) \quad \text{as } n \to \infty. \tag{40}$$

The sandwich variance estimator of $\widehat{\theta}$ is given by

$$\widehat{Var}(\widehat{\theta}) = \frac{1}{n} A_n(\widehat{\theta})^{-1} B_n(\widehat{\theta}) A_n(\widehat{\theta})^{-1\top}, \tag{41}$$

where

$$A_n(\widehat{\theta}) = -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta^\top} \Psi_{dr}(Y_i^*, T_i, X_i; \theta)|_{\theta=\widehat{\theta}} \tag{42}$$

and

$$B_n(\widehat{\theta}) = \frac{1}{n} \sum_{i=1}^{n} \Psi_{dr}(Y_i^*, T_i, X_i; \theta) \Psi_{dr}^\top(Y_i^*, T_i, X_i; \theta)|_{\theta=\widehat{\theta}} . \tag{43}$$

Then $\widehat{Var}(\widehat{\tau}_{\mathrm{DR}})$ is the element of the last row and the last column of $\widehat{Var}(\widehat{\theta})$.

Finally, we comment that Scenario 2 allows for different covariates-outcome associations for the treated and untreated groups and provides more flexibility than Scenario 1. However, implementing Scenario 2 involves separately estimating parameters of the outcome model for the treated and untreated groups. When one group has a small size, estimation results may be unsatisfactory. In this case, imposing common covariate effects for the treated and untreated groups as in Scenario 1 can help achieve reasonable estimation results.

Implementation in R and Examples

We develop an R package **ipwErrorY**, which implements the methods (Shu and Yi, 2017) described in the previous section. The developed package imports R packages **stats** and **nleqslv** (Hasselman, 2016). To illustrate the use of **ipwErrorY**, for each method, we simulate a dataset and then apply a function to analyze the dataset. To make sure users can reproduce the results, we use the function set.seed to generate data. Moreover, the simulated data provide users a clear sense about the data structure.

Implementation and example with known error

The function KnownError produces the ATE estimate using the correction method with known misclassification probabilities along with the sandwich-variance-based standard error and $(1 - \alpha)100\%$ confidence interval. Specifically, KnownError is defined as

```
KnownError(data, indA, indYerror, indX, sensitivity, specificity, confidence=0.95)
```

with arguments described in detail in **ipwErrorY** documentation. Below is an example to illustrate the use of KnownError.

We first load the package in R:

```
R> library("ipwErrorY")
```

Using sensitivity 0.95 and specificity 0.85, we create da, a dataset of size 2000 with "X1", "A" and "Yast" being the column names for the covariate, treatment and misclassified outcome, respectively:

```
R> set.seed(100)
R> X1 = rnorm(2000)
R> A = rbinom(2000, 1, 1/(1 + exp(-0.2 - X1)))
R> Y = rbinom(2000, 1, 1/(1 + exp(-0.2 - A - X1)))
R> y1 = which(Y == 1)
R> y0 = which(Y == 0)
R> Yast = Y
R> Yast[y1] = rbinom(length(y1), 1, 0.95)
R> Yast[y0] = rbinom(length(y0), 1, 0.15)
R> da = data.frame(X1 = X1, A = A,Yast = Yast)
```

By using the function head, we print the first six observations of dataset da so that the data structure is clearly shown as follows:

```
R> head(da)
          X1 A Yast
1 -0.50219235 1    1
2  0.13153117 1    1
3 -0.07891709 1    1
4  0.88678481 0    1
```

```
5  0.11697127 1    1
6  0.31863009 1    1
```

We call the developed function `KnownError` with sensitivity 0.95 and specificity 0.85, and obtain a list of the estimate, the sandwich-variance-based standard error and a 95% confidence interval:

```
R> KnownError(data = da, indA = "A", indYerror = "Yast", indX = "X1",
+     sensitivity = 0.95, specificity = 0.85, confidence=0.95)
$Estimate
[1] 0.1702513

$Std.Error
[1] 0.02944824

$`95% Confidence Interval`
[1] 0.1125338 0.2279688
```

Implementation and example with validation data

The function `EstValidation` produces the results for the method with validation data; they include the optimal linear combination estimate, the sandwich-variance-based standard error, $(1 - \alpha)100\%$ confidence interval and the estimated sensitivity and specificity. Specifically, `EstValidation` is defined as

```
EstValidation(maindata, validationdata, indA, indYerror, indX, indY, confidence=0.95)
```

with arguments described in detail in **ipwErrorY** documentation. Below is an example to illustrate the use of `EstValidation`.

Using sensitivity 0.95 and specificity 0.85, we create `mainda` which is the non-validation main data of size 1200, and `validationda` which is the validation data of size 800:

```
R> set.seed(100)
R> X1= rnorm(1200)
R> A = rbinom(1200, 1, 1/(1 + exp(-0.2 - X1)))
R> Y= rbinom(1200, 1, 1/(1 + exp(-0.2 - A - X1)))
R> y1 = which(Y == 1)
R> y0 = which(Y==0)
R> Yast = Y
R> Yast[y1] = rbinom(length(y1), 1, 0.95)
R> Yast[y0] = rbinom(length(y0), 1, 0.15)
R> mainda = data.frame(A = A, X1 = X1, Yast = Yast)
R> X1 = rnorm(800)
R> A = rbinom(800, 1, 1/(1 + exp(-0.2 - X1)))
R> Y = rbinom(800, 1, 1/(1 + exp(-0.2 - A - X1)))
R> y1 = which(Y == 1)
R> y0 = which(Y == 0)
R> Yast = Y
R> Yast[y1] = rbinom(length(y1), 1, 0.95)
R> Yast[y0] = rbinom(length(y0), 1, 0.15)
R> validationda = data.frame(A = A, X1 = X1, Y = Y, Yast = Yast)
```

We print the first six observations of non-validation data `mainda` and validation data `validationda`:

```
R> head(mainda)
  A          X1 Yast
1 1 -0.50219235   0
2 0  0.13153117   0
3 1 -0.07891709   1
4 1  0.88678481   1
5 0  0.11697127   1
6 1  0.31863009   1
R> head(validationda)
  A           X1 Y Yast
1 0 -0.0749961081 0    0
2 1 -0.9470827924 1    1
3 1  0.0003758095 1    1
```

```
4 0 -1.5249574007 0    0
5 1  0.0983516474 0    0
6 0 -1.5266078213 1    1
```

The preceding output clearly reveals that the non-validation data and validation data differ in the data structure. The non-validation data `mainda` record measurements of the treatment, covariate and misclassified outcome, indicated by the column names "A", "X1" and "Yast", respectively. In comparison, the validation data `validationda` record measurements of the treatment, covariate, misclassified outcome and the true outcome, indicated by the column names "A", "X1", "Yast", and "Y", respectively.

To apply the optimal linear combination method with validation data, we call the developed function `EstValidation` and obtain a list of the estimate, the sandwich-variance-based standard error, a 95% confidence interval, and the estimated sensitivity and specificity:

```
R> EstValidation(maindata = mainda, validationdata = validationda, indA = "A",
+    indYerror = "Yast",  indX = "X1",  indY = "Y", confidence=0.95)
$Estimate
[1] 0.1714068

$Std.Error
[1] 0.02714957

$`95% Confidence Interval`
[1] 0.1181946 0.2246189

$`estimated sensitivity and estimated specificity`
[1] 0.9482072 0.8557047
```

Implementation and example with replicates

The function `Est2Replicates` produces the results for the method with replicates; they include the estimate, the sandwich-variance-based standard error, $(1 - \alpha)100\%$ confidence interval, and the imposed constraint(s), and the information on sensitivity and specificity. Specifically, `Est2Replicates` is defined as

```
Est2Replicates(data, indA, indYerror, indX, constraint=c("sensitivity equals
  specificity", "known sensitivity", "known specificity", "known prevalence"),
  sensitivity = NULL, specificity = NULL, prevalence = NULL, confidence=0.95)
```

with arguments described in detail in **ipwErrorY** documentation. Below is an example to illustrate the use of `Est2Replicates`.

Using sensitivity 0.95 and specificity 0.85, we create da, a dataset of size 2000 with "A", "X1", and {"Yast1", "Yast2"} being the column names for the treatment, covariate, and two replicates of outcome, respectively:

```
R> set.seed(100)
R> X1 = rnorm(2000)
R> A = rbinom(2000, 1, 1/(1 + exp(-0.2 - X1)))
R> Y = rbinom(2000, 1, 1/(1 + exp(-0.2 - A - X1)))
R> y1 = which(Y == 1)
R> y0 = which(Y == 0)
R> Yast1 = Y
R> Yast1[y1] = rbinom(length(y1), 1, 0.95)
R> Yast1[y0] = rbinom(length(y0), 1, 0.15)
R> Yast2 = Y
R> Yast2[y1] = rbinom(length(y1), 1, 0.95)
R> Yast2[y0] = rbinom(length(y0), 1, 0.15)
R> da = data.frame(A = A, X1 = X1, Yast1 = Yast1, Yast2 = Yast2)
```

By using the function head, we print the first six observations of dataset da so that the data structure is clearly shown as follows:

```
R> head(da)
  A          X1 Yast1 Yast2
1 1 -0.50219235     1     1
2 1  0.13153117     1     1
```

```
3 1 -0.07891709    1    1
4 0  0.88678481    1    0
5 1  0.11697127    1    1
6 1  0.31863009    1    1
```

To apply the method with replicates, we call the developed function Est2Replicates with the imposed constraint that specificity equals 0.85. The following list of the estimate, the sandwich-variance-based standard error, a 95% confidence interval, the imposed constraint and the information on sensitivity and specificity is returned:

```
R> Est2Replicates(data = da,  indA = "A", indYerror = c("Yast1", "Yast2"),
+     indX = "X1", constraint = "known specificity", sensitivity = NULL,
+     specificity = 0.85, prevalence = NULL, confidence=0.95)
$Estimate
[1] 0.1908935

$Std.Error
[1] 0.02687287

$`95% Confidence Interval`
[1] 0.1382236 0.2435634

$`imposed constraint`
[1] "known specificity"

$`estimated sensitivity and assumed specificity`
[1] 0.95 0.85
```

Implementation and example of doubly robust estimation

The function KnownErrorDR produces the ATE estimate using the doubly robust correction method along with the sandwich-variance-based standard error and $(1 - \alpha)100\%$ confidence interval. Specifically, KnownErrorDR is defined as

```
KnownErrorDR(data, indA, indYerror, indXtrt, indXout, sensitivity, specificity,
  sharePara=FALSE, confidence=0.95)
```

with arguments described in detail in **ipwErrorY** documentation. Below is an example to illustrate the use of KnownErrorDR.

Using sensitivity 0.95 and specificity 0.85, we create da, a dataset of size 2000 with "A", {"X", "xx"} and "Yast" being the column names for the treatment, covariates and misclassified outcome, respectively:

```
R> set.seed(100)
R> X = rnorm(2000)
R> xx = X^2
R> A = rbinom(2000, 1, 1/(1 + exp(-0.1 - X - 0.2*xx)))
R> Y = rbinom(2000, 1, 1/(1 + exp(1 - A - 0.5*X - xx)))
R> y1 = which(Y == 1)
R> y0 = which(Y == 0)
R> Y[y1] = rbinom(length(y1), 1, 0.95)
R> Y[y0] = rbinom(length(y0), 1, 0.15)
R> Yast = Y
R> da = data.frame(A = A, X = X, xx = xx, Yast = Yast)
```

By using the function head, we print the first six observations of dataset da so that the data structure is clearly shown as follows:

```
R> head(da)
  A          X          xx Yast
1 1 -0.50219235 0.252197157    1
2 1  0.13153117 0.017300447    1
3 1 -0.07891709 0.006227907    1
4 0  0.88678481 0.786387298    0
5 1  0.11697127 0.013682278    1
6 1  0.31863009 0.101525133    0
```

When applying the doubly robust method with sensitivity 0.95 and specificity 0.85, covariates indicated by column names "X" and "xx" are both included in the treatment model and the outcome model. Let the outcome model be fit for the treated and untreated groups separately. We call the developed function KnownErrorDR and obtain a list of the estimate, the sandwich-variance-based standard error, and a 95% confidence interval:

```
R> KnownErrorDR(data = da, indA = "A", indYerror = "Yast", indXtrt = c("X", "xx"),
+     indXout = c("X", "xx"), sensitivity = 0.95, specificity = 0.85,
+     sharePara = FALSE, confidence=0.95)
$Estimate
[1] 0.2099162

$Std.Error
[1] 0.02811472

$`95% Confidence Interval`
[1] 0.1548124 0.2650201
```

Discussion

Misclassified binary outcome data arise frequently in practice and present a challenge in conducting causal inference. Discussion on addressing this issue is rather limited in the literature. Shu and Yi (2017) developed the IPW estimation methods for ATE with mismeasured outcome effects incorporated. To expedite the application of these correction methods, we develop an R package **ipwErrorY**. For practical settings where the treatment model and the outcome model are specified as logistic regression models, we implement the correction methods developed by Shu and Yi (2017) for settings with known misclassification probabilities, validation data, or replicates of the outcome data as well as the doubly robust method with known misclassification probabilities. Our package offers a useful and convenient tool for data analysts to perform valid inference about ATE when the binary outcome variable is subject to misclassification.

For each function of **ipwErrorY**, we implement the sandwich variance estimate to construct a normality-based confidence interval. Confidence intervals can also be constructed by bootstrapping (Efron, 1982; Efron and Tibshirani, 1993), which can be done by leveraging available functions of **ipwErrorY**. Below we provide example code to produce normality-based and percentile-based bootstrap confidence intervals for a doubly robust estimate with 200 bootstrap replicates.

```
R> drFUN<-function(dt) {
+   KnownErrorDR(data = dt, indA = "A", indYerror = "Yast", indXtrt = c("X", "xx"),
+             indXout = c("X", "xx"), sensitivity = 0.95, specificity = 0.85,
+             sharePara = FALSE, confidence=0.95)$`Estimate`
+ }
R> EST=drFUN(dt=da)
R> set.seed(100)
R> resultsBoot=replicate(200,drFUN(dt=da[sample(1:nrow(da),replace=TRUE),]))
R> STD=sd(resultsBoot)
R> lowN=EST-qnorm(1-(1-0.95)/2)*STD
R> upN=EST+qnorm(1-(1-0.95)/2)*STD
R> CIn=c(lowN,upN)
R> lowP=as.numeric(quantile(resultsBoot,probs=0.025))
R> upP=as.numeric(quantile(resultsBoot,probs=0.975))
R> CIp=c(lowP,upP)
```

We print the resultant bootstrap normality-based and percentile-based confidence intervals, respectively, as follows.

```
R> CIn
[1] 0.1562355 0.2635969
R> CIp
[1] 0.1610038 0.2655065
```

To make sure the users can reproduce the results, here we call the function set.seed before KnownErrorDR. If set.seed is not used, then the variance estimates generated at different times can differ due to the inner randomness of the bootstrap method.

This example code can be easily modified to produce bootstrap confidence intervals for an estimate obtained from a different method; one needs only to replace KnownErrorDR with the function in

ipwErrorY that corresponds to the method.

Package **ipwErrorY** requires the data be complete (i.e., no missing values). An error message is shown when NAs in the dataset are detected. For example, if we artificially introduce an NA in dataset da and call the developed function KnownErrorDR, an error message is displayed:

```
R> da[1,1]=NA
R> KnownErrorDR(data = da, indA = "A",  indYerror = "Yast", indXtrt = c("X", "xx"),
+              indXout = c("X", "xx"), sensitivity = 0.95, specificity = 0.85,
+              sharePara = FALSE, confidence=0.95)
Error in KnownErrorDR(data = da, indA = "A", indYerror = "Yast", indXtrt = c("X",  :
  invalid dataset with NAs (missing data detected)
```

Once seeing this error message, users need to check their dataset to see if the NAs can be replaced with suitable values. If missing values do occur, the easiest way is to take the subsample of complete observations to conduct analysis. The resulting point estimates can be reasonable if the missing data mechanism is missing completely at random (MCAR) (Little and Rubin, 2002); in this instance, efficiency loss can occur. However, when missing data are not MCAR, this procedure often yields biased results.

Our implementation uses a logistic regression model with a linear function of covariates for both the treatment and the outcome processes, as it is perhaps the most widely-used parametric tool to model a binary variable. Such a logistic regression model can be generalized to include additional terms, such as higher order terms, nonlinear functions, or interactions of the covariates. In this case, the users need only to first create an expanded dataset with those terms included as additional columns of new "covariates", and then use the **ipwErrorY** package to analyze the expanded dataset.

Acknowledgments

Bibliography

M. Babanezhad, S. Vansteelandt, and E. Goetghebeur. Comparison of causal effect estimators under exposure misclassification. *Journal of Statistical Planning and Inference*, 140:1306–1319, 2010. URL https://doi.org/10.1016/j.jspi.2009.11.015. [p1]

H. Bang and J. M. Robins. Doubly robust estimation in missing data and causal inference models. *Biometrics*, 61:962–973, 2005. URL https://doi.org/10.1111/j.1541-0420.2005.00377.x. [p1]

D. Braun, C. Zigler, F. Dominici, and M. Gorfine. Using validation data to adjust the inverse probability weighting estimator for misclassified treatment. *Harvard University Biostatistics Working Paper Series. Working Paper 201*, pages 1–19, 2016. [p1]

J. P. Buonaccorsi. *Measurement Error: Models, Methods, and Applications*. Chapman & Hall/CRC, London, 2010. [p1]

R. J. Carroll, D. Ruppert, L. A. Stefanski, and C. M. Crainiceanu. *Measurement Error in Nonlinear Models: A Modern Perspective, 2nd Edition*. Chapman & Hall/CRC, Boca Raton, 2006. [p1]

B. Efron. *The Jackknife, the Bootstrap and Other Resampling Plans*, volume 38. SIAM, Philadelphia, 1982. [p12]

B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. London: Chapman & Hall/CRC, 1993. [p12]

W. A. Fuller. *Measurement Error Models*. John Wiley & Sons, 1987. [p1]

P. Gustafson. *Measurement Error and Misclassification in Statistics and Epidemiology*. Chapman & Hall/CRC, London, 2003. [p1]

B. Hasselman. *Nleqslv: Solve Systems of Nonlinear Equations*, 2016. URL https://CRAN.R-project.org/package=nleqslv. R package version 3.0. [p8]

M. A. Hernán and J. M. Robins. *Causal Inference*. Chapman & Hall/CRC, Boca Raton, forthcoming, 2019. [p1]

C. C. Heyde. *Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation*. Springer-Verlag, 1997. [p2, 7]

G. W. Imbens and D. B. Rubin. *Causal Inference in Statistics, Social, and Biomedical Sciences*. Cambridge University Press, New York, 2015. [p1]

R. J. Little and D. B. Rubin. *Statistical Analysis with Missing Data, 2nd Edition*. John Wiley & Sons, 2002. [p13]

J. K. Lunceford and M. Davidian. Stratification and weighting via the propensity score in estimation of causal treatment effects: A comparative study. *Statistics in Medicine*, 23:2937–2960, 2004. URL https://doi.org/10.1002/sim.1903. [p1, 2]

D. F. McCaffrey, J. Lockwood, and C. M. Setodji. Inverse probability weighting with error-prone covariates. *Biometrika*, 100:671–680, 2013. URL https://doi.org/10.1093/biomet/ast022. [p1]

W. K. Newey and D. McFadden. Large sample estimation and hypothesis testing. *Handbook of Econometrics*, 4:2111–2245, 1994. [p2, 7]

J. M. Robins, A. Rotnitzky, and L. P. Zhao. Estimation of regression coefficients when some regressors are not always observed. *Journal of the American Statistical Association*, 89:846–866, 1994. URL https://doi.org/10.2307/2290910. [p1]

J. M. Robins, M. A. Hernán, and B. Brumback. Marginal structural models and causal inference in epidemiology. *Epidemiology*, 11:550–560, 2000. [p1]

P. R. Rosenbaum. Model-based direct adjustment. *Journal of the American Statistical Association*, 82: 387–394, 1987. URL https://doi.org/10.2307/2289440. [p1]

P. R. Rosenbaum. Propensity score. *In Encyclopedia of Biostatistics*, 5:3551–3555, 1998. [p1]

P. R. Rosenbaum and D. B. Rubin. The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55, 1983. URL https://doi.org/10.1093/biomet/70.1.41. [p1, 2]

K. J. Rothman, S. Greenland, and T. L. Lash. *Modern Epidemiology, 3rd Edition*. Lippincott Williams & Wilkins, Philadelphia, 2008. [p1]

D. O. Scharfstein, A. Rotnitzky, and J. M. Robins. Adjusting for nonignorable drop-out using semiparametric nonresponse models. *Journal of the American Statistical Association*, 94:1096–1120, 1999. URL https://doi.org/10.2307/2669930. [p1]

D. Shu and G. Y. Yi. Causal inference with measurement error in outcomes: Bias analysis and estimation methods. *Statistical Methods in Medical Research*, 2017. URL https://doi.org/10.1177/0962280217743777. [p1, 2, 3, 4, 5, 8, 12]

D. Shu and G. Y. Yi. *ipwErrorY: Inverse Probability Weighted Estimation of Average Treatment Effect with Misclassified Binary Outcome*, 2019. URL https://CRAN.R-project.org/package=ipwErrorY. R package version 2.1. [p1]

L. A. Stefanski and D. D. Boos. The calculus of m-estimation. *The American Statistician*, 56:29–38, 2002. URL https://doi.org/10.1198/000313002753631330. [p7]

I. White, C. Frost, and S. Tokunaga. Correcting for measurement error in binary and continuous variables using replicates. *Statistics in Medicine*, 20:3441–3457, 2001. URL https://doi.org/10.1002/sim.908. [p5]

G. Y. Yi. *Statistical Analysis with Measurement Error or Misclassification: Strategy, Method and Application*. Springer-Verlag, 2017. [p1, 2, 7]

G. Y. Yi and W. He. Analysis of case-ontrol data with interacting misclassified covariates. *Journal of Statistical Distributions and Applications*, 4:1–16, 2017. URL https://doi.org/10.1186/s40488-017-0069-0. [p5]

Di Shu
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo N2L 3G1, Ontario, Canada
shudi1991@gmail.com

Grace Y. Yi
Department of Statistics and Actuarial Science
University of Waterloo
Waterloo N2L 3G1, Ontario, Canada
yyi@uwaterloo.ca