Logistic regression analysis of randomized response data with missing covariates

S.H. Hsieh\textsuperscript{a}, S.M. Lee\textsuperscript{b,*}, P.S. Shen\textsuperscript{a}

\textsuperscript{a}Department of Statistics, Tunghai University, Taiwan
\textsuperscript{b}Department of Statistics, Feng Chia University, Taiwan

\begin{abstract}
Randomized response is an interview technique designed to eliminate response bias when sensitive questions are asked. In this paper, we present a logistic regression model on randomized response data when the covariates on some subjects are missing at random. In particular, we propose Horvitz and Thompson (1952)-type weighted estimators by using different estimates of the selection probabilities. We present large sample theory for the proposed estimators and show that they are more efficient than the estimator using the true selection probabilities. Simulation results support theoretical analysis. We also illustrate the approach using data from a survey of cable TV.
\end{abstract}

\section{Introduction}

Most people are reluctant to answer questions about sensitive topics publicly, like drug or alcohol (ab)use, sexuality or anti-social behavior. As a result, respondents may refuse to give the embarrassing answer and the stigmatizing behavior is often underreported. The collection of data through personal interview surveys on sensitive questions is a serious issue. Randomized response technique (RRT) is an interview that was designed to protect the privacy of the respondent. Using two complementary sensitive questions, Warner (1965) introduced a related-question design, for instance:

Q1: I have cheated on an examination by copying answers.
Q2: I have never cheated on an examination by copying answers.

A chance game (for instance with dice or coins) is used to decide which of the two questions is answered with “Yes” or “No”. In RRT, the answer to a sensitive question depends partly on the respondent’s true status and partly on the outcome of a randomizing device. Similar to Warner’s RRT design, Horvitz et al. (1967) and Greenberg et al. (1969) proposed an unrelated question design, where one question is sensitive, whereas the second question is nonsensitive or innocuous, for example:

Q1: I have cheated on an examination by copying answers.
Q2*: I was born in January, February, or March.

Since such techniques do not reveal to the interviewer the group to which a respondent belongs, it is expected that we can get an accurate estimate of the true prevalence in the population of the attitudes towards an examination by copying answers. More complex RRT designs were proposed by Boruch (1971) and Kuk (1990).
Maddala (1983) was the first to discuss logistic regression model for Warner’s RRT data with completely observable covariates. Scheers and Dayton (1988) presented theory for a covariate randomized response model that is an extension of the Warner (1965) procedure and for a covariate extension of the unrelated-question RRT (Greenberg et al., 1969). Corstange (2004) proposed a method to estimate parameters in a hidden logistic regression. Recently, there have been some researches regarding multiple RRT variables (e.g. Böckenholt and van der Heijden, 2007; Fox, 2005; van den Hout et al., 2007).

In many RRT applications, the RRT variables and covariates may not be completely observable. In this paper, we shall demonstrate how to analyze data from the unrelated-question RRT (Greenberg et al., 1969) with missing covariates. Let \( Y \) be a latent binary RRT variable corresponding to an answer to the innocuous question, \( Z \) be a vector of covariates which is always observed, \( X \) be another covariate vector that may be missing on some subjects, and \( W \) be a surrogate variable for \( X \) and independent of \( Y \) given \( (X,Z) \). Let \( Y = 1 \) denote the event of answering “Yes”, and \( Y = 0 \) denote the event of answering “No”. We consider the following logistic regression model:

\[
P(Y = 1|X,Z,W) = H(\beta_0 + \beta_1^TX + \beta_2^TZ) = H(\beta^TX),
\]

where \( H(u) = (1 + \exp(-u))^{-1} \), \( X' = (1, X'Z)' \) and \( \beta = (\beta_0, \beta_1', \beta_2')' \) is a vector of unknown parameters. When \( Y \) is observable and \( X \) is missing on some subject, several methods have been proposed for estimating regression parameters \( \beta \) of model (1). For example, Breslow and Cain (1988) proposed a pseudo-conditional likelihood method, Reilly and Pepe (1995) proposed a mean-score estimator, and Wang et al. (2002) proposed a joint conditional likelihood estimator by combining validation data set and nonvalidation data set. When missingness does not depend on either \( Y \) or \( X \), Carroll and Wand (1991) and Pepe and Fleming (1991) proposed a likelihood estimator of \( \beta \). When missingness does not depend on \( X \) (i.e. missing at random (MAR), see Rubin, 1976), Flander and Greenland (1991) and Zhao and Lipsitz (1992) suggested a weighted estimator of \( \beta \).

In Section 2, under model (1), we describe the unrelated-question RRT design. In Section 3, when \( X \) is missing at random (MAR, Rubin, 1976), we propose Horvitz and Thompson (1952)-type (HT-type) weighted estimators. We focus primarily on efficiency comparisons among HT-type weighted estimators by using different estimation of selection probabilities. In Section 4, we conduct a simulation study to investigate the performance of the proposed estimators. In Section 5, the proposed estimators are applied to a survey of cable TV. In Section 7, we provide some concluding remarks.

### 2. The unrelated-question RRT design

In this section, we start with the unrelated-question RRT (Horvitz et al., 1967; Greenberg et al., 1969) as an example and investigate the relationship between unrelated-question RRT model and related-question RRT model considered by Warner’s (1965).

Let \( T \) be a latent binary RRT variable corresponding to an answer to the innocuous question, \( Y^0 \) be the binary variable of the observed answer, i.e. the response of sensitive question (\( Y \)) or innocuous question (\( T \)), and \( S \) be a latent binary variable based on a chance game. Therefore, we have \( Y^0 = Y \) with probability \( P(S = 1) = p \) and \( Y^0 = T \) with probability \( P(S = 0) = 1 - p \). Note that \( p \) is the known probability of answering sensitive question. Based on model (1), the following diagram shows how the observed \( Y^0 \) variable relates to covariates \( X \) and \( Z \) (see Fig. 1).

Conditional on \( (X,Z,W) \), the probability of answering “YES” is

\[
P(Y^0 = 1|X,Z,W) = P(Y^0 = 1|S = 1,X,Z,W)P(S = 1) + P(Y^0 = 1|S = 0,X,Z,W)P(S = 0)
= P(Y^0 = 1) = H_2(\beta^TX)
= H(\beta^TX)p + k,
\]

where \( k = c(1 - p) \).

![Fig. 1. The unrelated RRT system.](image-url)
In the original related-question RRT by Warner (1965), we have \( P(T = 1|X,Z,W) = P(Y = 0|X,Z,W) = 1 - H(\beta^T X) \). Therefore, the probability of answering "Yes" is as follows:

\[
P(Y^0 = 1|X,Z,W) = H(\beta^T X)p + [1 - H(\beta^T X)] \times (1 - p)
\]

\[
= (2p - 1)H(\beta^T X) + (1 - p)
\]

\[
= p'H(\beta^T X) + \frac{1 - p'}{2},
\]

where \( p \in (\frac{1}{2}, 1) \) and \( p' = 2p - 1 \). Hence if we write \( c = \frac{1}{2} \), the related-question RRT model can be described in the same way as unrelated-question RRT model. In this paper, we assume that both \( p \) and \( c \) are known. In practice, this is a reasonable assumption since we can obtain \( p \) and \( c \) from the applicable design.

3. The proposed estimators

Let \( n \) denotes the sample size. When \( X \) is missing on some subject, for \( i = 1, \ldots, n \), we can only observe \((Y_i^0, X_i, Z_i, W_i)\) or \((Y_i^0, Z_i, W_i)\). Let \( V_i = (Z_i^T, W_i^T) \) and \( \delta_i \) indicate whether \( X_i \) is observed \((\delta_i = 1)\) or not \((\delta_i = 0)\). The validation data set \((\delta_i = 1)\) consists of \((Y_i^0, X_i, V_i)\) and nonvalidation data set \((\delta_i = 0)\) consists of \((Y_i^0, V_i)\). Assume that \( X_i \) is missing at random (MAR, see Rubin, 1976), the selection probability depends on \((Y_i^0, V_i)\) but not on \( X_i \), i.e. \( P(\delta_i = 1|Y_i^0, V_i) = P(Y_i^0, V_i) \), which is assumed to be strictly positive.

For the unrelated-question RRT data without missing covariate \( X_i \), under model (2), the likelihood is given by (see Scheers and Dayton, 1988)

\[
L = \prod_{i=1}^{n} \left( P(Y_i^0 = 1|X_i, Z_i, W_i) \right)^{Y_i^0} \left( P(Y_i^0 = 0|X_i, Z_i, W_i) \right)^{1-Y_i^0} = \prod_{i=1}^{n} \left( H(\beta^T X)p + k \right)^{Y_i^0} \left( 1 - H(\beta^T X)p - k \right)^{1-Y_i^0}.
\]

Hence, the maximum likelihood estimator (MLE) \( \hat{\beta} \) of \( \beta \) can be obtained by solving the following estimating equation:

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \delta_i X_i A_i(\beta) (Y_i^0 - pH(\beta^T X_i) - k) = 0,
\]

where \( A_i(\beta) = pH(\beta^T X_i)/[pH(\beta^T X_i) + k][1 - pH(\beta^T X_i) - k] \) and \( H(\beta^T X_i) = H(\beta^T X_i)[1 - H(\beta^T X_i)] \).

When \( Y_i \) is observed and \( X_i \) is MAR, Flander and Greenland (1991) and Zhao and Lipsitz (1992) proposed the Horvitz and Thompson (1952)-type (HT-type) weighted estimators. Under unrelated-question RRT, the HT-type weighted estimating equation \( U_n(\beta, \pi) \) is defined as follows:

\[
U_n(\beta, \pi) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\delta_i}{\pi(Y_i^0, V_i)} X_i A_i(\beta) (Y_i^0 - pH(\beta^T X_i) - k).
\]

When \( \pi(Y_i^0, V_i) \)'s are known, by direct calculation we have

\[
E \left\{ \frac{\delta_i}{\pi(Y_i^0, V_i)} X_i A_i(\beta) (Y_i^0 - pH(\beta^T X_i) - k) \right\} = 0.
\]

Therefore, the estimating equation (4) is unbiased. Let \( \hat{\beta}_1 \) denote the solution of \( U_n(\beta, \pi) = 0 \). It is easy to see that when data are complete (i.e. \( \pi(Y_i^0, V_i) = 1 \), for any \( Y_i^0 \) and \( V_i \)), the estimators \( \hat{\beta} \) and \( \hat{\beta}_1 \) are equivalent to each other.

Similar to the approach of Zhao and Lipsitz (1992), it can be shown that the estimator \( \hat{\beta}_1 \) is consistent and \( \sqrt{n} (\hat{\beta}_1 - \beta) \) is asymptotically normal provided that the probabilities \( \pi(Y_i^0, V_i) \)'s are correctly specified. In practice, the selection probabilities \( \pi(Y_i^0, V_i) \)'s are usually unknown. Hence, we need to estimate nuisance parameters \( \pi(Y_i^0, V_i) \)'s for the estimation of \( \beta \). Based parametric and nonparametric estimators of \( \pi(Y_i^0, V_i) \)'s, we shall propose two estimators of \( \beta \).

First, we consider a parametric estimator of \( \pi_i(\mathbf{z}) \) using the following logistic regression model of \( \delta_i \) on \((Y_i^0, V_i)\):

\[
\pi_i(\mathbf{z}) = P(\delta_i = 1|Y_i^0, V_i) = H(z_0 + z_1 Y_i^0 + z_2 V_i) = H(\mathbf{z}^T V_i),
\]

where \( V_i = (1, Y_i^0, V_i) \) and \( \mathbf{z} = (z_0, z_1, z_2) \) which are some unknown parameters.

When \( \mathbf{z} \) is known, it can be shown by direct calculation that \( E(\delta_i/\pi_i(\mathbf{z})) X_i A_i(\beta) (Y_i^0 - pH(\beta^T X_i) - k) = 0 \). Define an estimating score \( U_n(\beta, \pi) \) as follows:

\[
U_n(\beta, \pi) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \delta_i \pi_i(\mathbf{z}) X_i A_i(\beta) (Y_i^0 - pH(\beta^T X_i) - k) \right\}.
\]

The weighted estimator with known \( \pi_i(\mathbf{z}) \), denoted by \( \hat{\beta}_1^* \), is obtained by solving \( U_n(\beta, \pi) = 0 \). By (4) and (6), it follows that \( \hat{\beta}_1^* \) and \( \hat{\beta}_1 \) are equivalent to each other.
When $\alpha$ is unknown, we need to estimate $\alpha$ before solving the score equation $U_0(\beta, \alpha)$. The MLE $\hat{\alpha}$ of $\alpha$ can be obtained by solving $1/\sqrt{n}\sum_{i=1}^{n} \delta_i I(Y_i^0 = y_i, V_j = v_j) = 0$. Since the estimating equation is unbiased, $\hat{\alpha}$ is a consistent estimator of $\alpha$. The estimating equation $U_n(\beta, \hat{\alpha})$ is obtained by replacing $\hat{\alpha}$ in (6) by $\alpha$. Hence, the weighted parametric estimator of $\beta$, denoted by $\hat{\beta}_{wp}$, is obtained by solving $U_n(\beta, \hat{\alpha}) = 0$. Note that this approach yields consistent estimators of the regression coefficients of model (2), provided that $\pi_i(\alpha)$ is correctly specified. However, misspecification of $\pi_i(\alpha)$ may lead to biased estimators of $\beta$.

Next, we consider a nonparametric model for the selection probability. Let $v_1, \ldots, v_m$ denote the distinct values of the $V_j$'s. For $v \in (v_1, \ldots, v_m)$ and $y^0 = 0, 1$, we consider a nonparametric estimator of $\pi(Y_i^0, V_j)$ as follows:

$$
\hat{\pi}(y_i^0, v) = \frac{\sum_{j=1}^{m} \delta_i I(Y_i^0 = y_i^0, V_j = v) \pi(Y_i^0, V_j)}{\sum_{j=1}^{m} \pi(Y_i^0, V_j)}.
$$

(7)

where $I(\cdot)$ is an indicate function. Based on (7), the weighted semiparametric estimator of $\beta$, denoted by $\hat{\beta}_{ws}$, can be obtained by solving

$$
U_n(\beta, \hat{\pi}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{\delta_i}{\hat{\pi}(Y_i^0, V_i)} A_i(\beta; Y_i^0, V_i - pH(\beta^T X_i) - k) \right\} = 0.
$$

(8)

4. Asymptotic theory

In this section, we provide asymptotic results under the assumption that $V$ is discrete and $X$ is MAR. The following regularity conditions are crucial in deriving the asymptotic properties of $\hat{\beta}_t$, $\hat{\beta}_{us}$, and $\hat{\beta}_{wp}$.

(A1) Let $\text{supp}(V)$ denote the support of $V$. For any $y^0 = 0, 1$ and $v \in \text{supp}(V)$, the selection probability $\pi(y^0, v) > 0$.

(A2) $\mathbb{E}(XY^T A_t(\beta; H(\beta^T X))]$ is positive definite in a neighborhood of the true $\beta$, where $H(\beta^T X) = H(\beta^T X)[1 - H(\beta^T X)]$.

(A3) $S_t(\beta) = X_i A_i(\beta; \beta_0, Y_i^0 - pH(\beta^T X) - k)$ has a finite second moment definite in a neighborhood of the true $\beta$.

(A4) The first derivatives of $U_t(\cdot)$ with respect to $\beta$ exist almost surely in a neighborhood of the true $\beta$. Further, in such a neighborhood, the second derivatives are bounded above by a function of $(Y_i^0, X, V)$, whose expectation exists.

(A5) $\mathbb{E}(V^T \pi(\alpha)[1 - \pi(\alpha)])$ is positive definite in a neighborhood of the true $\alpha$.

The following lemma derives the asymptotic properties of $\hat{\beta}_t$. For any nonsingular matrix $G$, define $G^{-T} = [G^{-1}]^T$. Let $S_t(\beta) = X_i A_i(\beta; \beta_0, Y_i^0 - pH(\beta^T X) - k)$.

Lemma 1. Under the conditions (A1)-(A3) and when the $\pi(Y_i^0, V_j)$’s are known, $\hat{\beta}_t$ is a consistent estimator of $\beta$ and $\sqrt{n}(\hat{\beta}_t - \beta)$ has an asymptotic normal distribution with mean 0 and covariance matrix

$$
A_t = G^{-1}([\beta, \pi])G^{-T}([\beta, \pi]),
$$

where

$$
G(\beta, \pi) = \mathbb{E}(\chi X^T pA(\beta; H(\beta^T X))),
$$

and

$$
J(\beta, \pi) = \mathbb{E} \left\{ S_t(\beta) [S_t(\beta)]^T \right\}.
$$

The proof of the Lemma 1 is deferred to Appendix. Next, we derive a consistent estimator of $A_t$. Let

$$
G_n(\hat{\beta}_t, \pi) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(Y_i^0, V_i)} X_i A_i(\beta; Y_i^0, V_i - pH(\hat{\beta}_t^T X_i) - k),
$$

and

$$
J_n(\hat{\beta}_t, \pi) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{\hat{\pi}(Y_i^0, V_i)} X_i A_i^2(\beta; Y_i^0, V_i - pH(\hat{\beta}_t^T X_i) - k)^2.
$$

Since $\hat{\beta}_t \overset{P}{\rightarrow} \beta$, we have $G_n(\hat{\beta}_t, \pi) \overset{P}{\rightarrow} G(\beta, \pi)$ and $J_n(\hat{\beta}_t, \pi) \overset{P}{\rightarrow} J(\beta, \pi)$, where $\overset{P}{\rightarrow}$ denotes convergence in probability. Hence a consistent estimator of $A_t$ is given by

$$
\hat{A}_t = G_n^{-1}(\hat{\beta}_t, \pi) J_n(\hat{\beta}_t, \pi) G_n^{-T}(\hat{\beta}_t, \pi).
$$
We now present the asymptotic properties of \( \hat{\beta}_{wp} \). Define
\[
G(\beta, \mathbf{x}) = E(\mathbf{X}^T \mathbf{pA(\beta)H}^{-1})(\beta^T, \mathbf{x}),
\]
\[
J(\beta, \mathbf{x}) = E\left\{ \frac{S_1(\beta)[S_1(\beta)]^T}{\pi(\mathbf{x})} \right\},
\]
\[
D(\beta, \mathbf{x}) = E[(1 - \pi(\mathbf{x})) \mathbf{Y}^T A(\beta)Y_0 - pH(\beta^T, \mathbf{x}) - k],
\]
and
\[
K(\mathbf{x}) = E(\mathbf{Y}^T \pi(\mathbf{x})(1 - \pi(\mathbf{x}))).
\]

**Theorem 1.** Under the conditions (A1)–(A5), \( \hat{\beta}_{wp} \) is a consistent estimator of \( \beta \) and \( \sqrt{n}(\hat{\beta}_{wp} - \beta) \) has an asymptotic normal distribution with mean 0 and covariance matrix
\[
A_{wp} = G^{-1}(\beta, \mathbf{x})(J(\beta, \mathbf{x}) - D(\beta, \mathbf{x})K^{-1}(\mathbf{x})D^T(\beta, \mathbf{x}))G^{-T}(\beta, \mathbf{x}).
\]

The proof of Theorem 1 is deferred to Appendix. We now derive a consistent estimator of \( A_{wp} \). Let
\[
G_n(\hat{\beta}_{wp}, \hat{\mathbf{x}}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{\pi(\mathbf{x}_i)} \mathbf{X}_i \mathbf{X}_i^T (pA_0(\hat{\beta}_{wp})H^{-1})(\hat{\beta}_{wp}, \mathbf{x}_i);
\]
\[
J_n(\hat{\beta}_{wp}, \hat{\mathbf{x}}) = \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{\pi(\mathbf{x}_i)} \mathbf{X}_i \mathbf{X}_i^T \frac{1}{\pi(\mathbf{x}_i)} S_1(\hat{\beta}_{wp})[\hat{\beta}_{wp}]^T Y_0^0 - pH(\hat{\beta}_{wp}, \mathbf{x}_i) - k^2,
\]
\[
D_n(\hat{\beta}_{wp}, \hat{\mathbf{x}}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1 - \pi(\mathbf{x}_i)}{\pi(\mathbf{x}_i)} \delta_i \mathbf{X}_i \mathbf{Y}_i [\hat{\beta}_{wp}, \mathbf{x}_i] - k - pH(\hat{\beta}_{wp}, \mathbf{x}_i) - k,
\]
and
\[
K_n(\hat{\mathbf{x}}) = n^{-1} \sum_{i=1}^{n} \mathbf{Y}_i \mathbf{Y}_i^T \pi(\mathbf{x}_i)(1 - \pi(\mathbf{x}_i)).
\]

The weak law of large number properties are summarized as follows: \( \hat{\beta}_{wp} \) \( \overset{p}{\rightarrow} \beta, \hat{\mathbf{x}} \) \( \overset{p}{\rightarrow} \mathbf{x}, G_n(\hat{\beta}_{wp}, \hat{\mathbf{x}}) \) \( \overset{p}{\rightarrow} G(\beta, \mathbf{x}), J_n(\hat{\beta}_{wp}, \hat{\mathbf{x}}) \) \( \overset{p}{\rightarrow} J(\beta, \mathbf{x}), K_n(\hat{\mathbf{x}}) \) \( \overset{p}{\rightarrow} K(\mathbf{x}), \) and \( D_n(\hat{\beta}_{wp}, \hat{\mathbf{x}}) \) \( \overset{p}{\rightarrow} D(\beta, \mathbf{x}). \) Hence a consistent estimator of \( A_{wp} \) is given by
\[
\hat{A}_{wp} = G_n^{-1}(\hat{\beta}_{wp}, \hat{\mathbf{x}})(J_n(\hat{\beta}_{wp}, \hat{\mathbf{x}}) - D_n(\hat{\beta}_{wp}, \hat{\mathbf{x}})K_n^{-1}(\hat{\mathbf{x}})D_n^T(\hat{\beta}_{wp}, \hat{\mathbf{x}}))G_n^{-T}(\hat{\beta}_{wp}, \hat{\mathbf{x}}).
\]

**Theorem 2.** Under conditions (A1)–(A4), \( \hat{\beta}_{ws} \) is a consistent estimator of \( \beta \) and \( \sqrt{n}(\hat{\beta}_{ws} - \beta) \) has an asymptotic normal distribution with mean 0 and covariance matrix
\[
A_{ws} = G^{-1}(\beta, \pi)(J(\beta, \pi) - [J^*(\beta, \pi) - C^*(\beta, \pi)])G^{-T}(\beta, \pi),
\]
where
\[
G(\beta, \pi) = E(\mathbf{X}^T \mathbf{pA(\beta)H}^{-1})(\beta^T, \mathbf{x}),
\]
\[
J(\beta, \pi) = E\left\{ \frac{S_1(\beta)[S_1(\beta)]^T}{\pi(\mathbf{Y}_i^0, \mathbf{V}_i)} \right\},
\]
\[
J^*(\beta, \pi) = E\left\{ \frac{S_1(\beta)[S_1(\beta)]^T}{\pi(\mathbf{Y}_i^0, \mathbf{V}_i)} \right\},
\]
\[
C^*(\beta, \pi) = E[S_1(\beta)[S_1(\beta)]^T] \quad \text{and} \quad S_1(\beta) = E[S_1(\beta)|\mathbf{Y}_i^0, \mathbf{V}_i].
\]

The proof of Theorem 2 is deferred to Appendix. We now give a consistent estimator of \( A_{ws} \). Let
\[
G_n(\hat{\beta}_{ws}, \hat{\pi}) = n^{-1} \sum_{i=1}^{n} \frac{\delta_i}{\pi(\mathbf{Y}_i^0, \mathbf{V}_i)} \mathbf{X}_i \mathbf{X}_i^T (pA_0(\hat{\beta}_{ws})H^{-1})(\hat{\beta}_{ws}, \mathbf{x}_i),
\]
and
\[
S_n(\hat{\beta}_{ws}) = \left\{ \sum_{i=1}^{n} \delta_i S_1(\hat{\beta}_{ws})|\mathbf{Y}_i^0 = \mathbf{Y}_i^0, \mathbf{V}_j = \mathbf{V}_j \right\} / \left\{ \sum_{i=1}^{n} \delta_i |\mathbf{Y}_i^0 = \mathbf{Y}_i^0, \mathbf{V}_j = \mathbf{V}_j \right\}.
\]
Since \( \hat{\beta}_{ws} \) \( \overset{p}{\rightarrow} \beta \) and \( \hat{\pi}(\mathbf{Y}_i^0, \mathbf{V}_j) \) \( \overset{p}{\rightarrow} \pi(\mathbf{Y}_i^0, \mathbf{V}_j) \), we have \( G_n(\hat{\beta}_{ws}, \hat{\pi}) = G(\beta, \pi). \) For any vector \( \mathbf{a} \), define \( \mathbf{a} \odot \mathbf{a}^T = \mathbf{a} \mathbf{a}^T \). Hence a consistent estimator of the \( A_{ws} \) is given by
\[
\hat{A}_{ws} = G_n^{-1}(\hat{\beta}_{ws}, \hat{\pi})\left\{ \frac{1}{n} \sum_{i=1}^{n} \frac{\delta_i}{\pi(\mathbf{Y}_i^0, \mathbf{V}_i)} S_1(\hat{\beta}_{ws}) + \left( 1 - \frac{\delta_i}{\pi(\mathbf{Y}_i^0, \mathbf{V}_i)} \right) S_1(\hat{\beta}_{ws}) \right\} \odot G_n^{T}(\hat{\beta}_{ws}, \hat{\pi}).
\]
Note that the average score \( S_n(\hat{\beta}_{ws}) \) is obtained from the validation set.
In the following, we analytically investigate the asymptotic variances of the proposed HT-type weighted estimators under different estimated selection probabilities. When \( \alpha \) in (5) is known, by Lemma 1 and replacing \( \pi_i(\alpha) \) by \( \pi(Y_i^0, V_i) \), we have

\[
A_1 = G^{-1}(\beta, \alpha)j(\beta, \alpha)G^{-T}(\beta, \alpha).
\]

Therefore, by Theorem 1, we have

\[
A_t - A_{\text{wp}} = G^{-1}(\beta, \alpha)j(D(\beta, \alpha)K^{-1}(\alpha)D^T(\beta, \alpha))G^{-T}(\beta, \alpha).
\]

(9)

Similarly, by Lemma 1 and Theorem 2, we have

\[
A_t - A_{\text{ws}} = G^{-1}(\beta, \pi)j^*(\beta, \pi) - C^*(\beta, \pi)G^{-T}(\beta, \pi).
\]

(10)

Then both matrices \( A_t - A_{\text{wp}} \) and \( A_t - A_{\text{ws}} \) in (9) and (10) are positive definite. Hence, the estimator using estimated selection probabilities in general leads to more efficient estimator compared to that using the true selection probabilities, i.e., the efficiency of estimator of \( \beta \) can be increased via data adjustment of \( \pi(Y_i^0, V_i) \) (see Robins et al., 1994; Rosenbaum, 1987). Since we cannot show the matrix \( A_{\text{ws}} - A_{\text{wp}} \) is positive definite, the asymptotic behavior of \( \beta_{\text{ws}} \) and \( \beta_{\text{wp}} \) are investigated through simulation study in the next section.

5. Simulation study

In this section, a simulation study is conducted to investigate the finite sample performance under various estimates of \( \pi(\cdot) \). Assume that \( p \) and \( c \) are known. The values of \( p \), \( c \), and sample size \( n \) are set as follows:

1. \( p = 0.5 \) and 0.7.
2. \( n = 500 \) and 1000.
3. \( c = 0.5 \) and 0.8.

For each combination of \( p \), \( c \), and \( n \), we calculate the following three estimators:

- \( \hat{\beta}_t \): the weighted estimator using true selection probabilities.
- \( \hat{\beta}_{\text{wp}} \): the weighted parametric estimator.
- \( \hat{\beta}_{\text{ws}} \): the weighted semiparametric estimator.

The replication is 1000 times. For each estimator, we compute bias, asymptotic standard error (ASE), standard deviation (SD), and the 95% confidence interval coverage probabilities (CP). The three cases are explored as follows:

Case 1:

We consider the case with one univariate covariate \( X \). First, the \( U \)'s and \( \varepsilon \)'s were generated from normal distributions \( N(0, 1) \) and \( N(0, \sigma) \), respectively. Given \( U \), a binary covariate was defined by \( X = I(U \geq 0) \). Given \( U \) and \( \varepsilon \), a binary surrogate covariate was defined by \( W = I(U + \varepsilon \geq 0) \). The binary outcome \( Y^0 \) was generated as follows: both \( Y \) and \( T \) were generated as a binary variable with \( P(Y = 1 | X) = H(\beta_0^0 + \beta_1^0 X) \), \( (\beta_0^0, \beta_1^0) = (0.5, \ln(2)) \) and \( P(T = 1 | X) = C \). Given \( Y \) and \( T \), the \( Y^0 \) was generated as \( P(Y^0 = Y) = p \) and \( P(Y^0 = T) = 1 - p \). The validation data indicator \( \delta \) was a binary variable with \( P(\delta = 1 | Y^0, W) = H(0.5 - 0.5Y^0 + 0.5W) \), which resulted in about 40% of missing rate (mr). For example, when \( p = 0.7 \), \( n = 500 \), and \( mr = 40\% \), there were 210 validation data of responding the question of interest.

Simulation results (see Table 1) show that the efficiencies of all the three estimators increase as \( c \) and \( n \) increases. Note that the probability \( P(Y^0 = 1 | X) \) increases as the value of \( k = c(1 - p) \) increases. For the estimation of \( \beta \), the estimator \( \hat{\beta}_{\text{ws}} \) outperforms both estimators \( \hat{\beta}_{\text{wp}} \) and \( \hat{\beta}_t \), and \( \hat{\beta}_{\text{wp}} \) performs slightly better than \( \hat{\beta}_t \). This result supports theoretical analysis in Section 4. Moreover, the correlation coefficients between \( X \) and \( W \), denoted by \( \text{corr}(X, W) \), were about 0.84 and 0.50 for \( \sigma = 0.25 \) and 1, respectively. Simulation results show that all the estimators perform better when \( \sigma = 0.25 \) compared to \( \sigma = 1 \). In this case, the efficiencies of the estimator \( \hat{\beta}_{\text{ws}} \) increase as \( \sigma \) decreases.

We also investigated the efficiencies of the three estimator \( \hat{\beta}_t \), \( \hat{\beta}_{\text{ws}} \), and \( \hat{\beta}_{\text{wp}} \) under various values of \( \text{corr}(X, W) \). We selected various \( \sigma \) values so that \( \text{corr}(X, W) \) ranged from 0.14 to 0.84. The ratio of the asymptotic variances (ASV) of \( \hat{\beta}_{\text{wp}} \) and \( \hat{\beta}_{\text{ws}} \) to that of \( \hat{\beta}_t \) are computed. Fig. 2 presents the result of the case \( n = 1000 \), \( p = 0.7 \) and \( c = 0.8 \), where \( \text{RE}_{1k} = \text{ASV}(\hat{\beta}_{\text{wp}})/\text{ASV}(\hat{\beta}_t) \), \( \text{RE}_{2k} = \text{ASV}(\hat{\beta}_{\text{ws}})/\text{ASV}(\hat{\beta}_t) \) and \( k = 0.1 \) for \( \beta_0^0 \) and \( \beta_1^0 \), respectively. Based on Fig. 1, we see that all \( \text{RE}_{1k} \) and \( \text{RE}_{2k} \) for \( k = 0.1 \) are smaller than one, which implies that both estimators \( \hat{\beta}_{\text{ws}} \) and \( \hat{\beta}_{\text{wp}} \) outperforms \( \hat{\beta}_t \). When \( W \) is highly informative about \( X \) (i.e., the \( \text{corr}(X, W) \) is large), the estimator \( \hat{\beta}_{\text{ws}} \) performs much better than \( \hat{\beta}_{\text{wp}} \). When \( \text{corr}(X, W) \) is not large, the estimator \( \hat{\beta}_{\text{ws}} \) still performs slightly better than \( \hat{\beta}_{\text{wp}} \). Note that the \( \text{RE}_{10} \) and \( \text{RE}_{11} \) are flat, which indicates that \( \text{corr}(X, W) \) have little impact on the relative efficiency of \( \hat{\beta}_{\text{wp}} \) to \( \hat{\beta}_t \). However, the relative efficiency of \( \hat{\beta}_{\text{ws}} \) to \( \hat{\beta}_t \) (i.e., \( \text{RE}_{20} \) and \( \text{RE}_{21} \)) increases as \( \text{corr}(X, W) \) increases.
Case 2:
The covariate $X$ was generated from a uniform distribution $U(-1, 1)$. The $e$’s was generated from a uniform distribution $U(-0.5, 0.5)$. Given $e$, a binary surrogate covariate was defined by $W = I(X + e \geq 0)$. The covariate $Z$ was a binary variable with $P(Z = 1) = P(Z = 0) = 0.5$. Given $X$ and $Z$, $Y$ was a binary variable with $P(Y = 1|X, Z) = H(\beta_0 + \beta_1 X + \beta_2 Z)$ where $\beta = (0.5, 1, -1)$. The distribution of $T$ and $Y^0$ are the same as those used in Case 1. Given $Y^0$ and $W$, the validation data

Table 1
Simulation results of Case 1 (with univariate covariate $X$).

<table>
<thead>
<tr>
<th>$c$</th>
<th>$n$</th>
<th>$\beta_0$ Bias</th>
<th>$\hat{\beta}_{wp}$</th>
<th>$\hat{\beta}_{ws}$</th>
<th>$\beta_1$ Bias</th>
<th>$\hat{\beta}_{wp}$</th>
<th>$\hat{\beta}_{ws}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8</td>
<td>500</td>
<td>0.0092</td>
<td>0.0090</td>
<td>0.0119</td>
<td>0.0048</td>
<td>0.0013</td>
<td>-0.0014</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2323</td>
<td>0.1992</td>
<td>0.1821</td>
<td>0.2371</td>
<td>0.2071</td>
<td>0.1977</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2388</td>
<td>0.2077</td>
<td>0.1887</td>
<td>0.2351</td>
<td>0.2068</td>
<td>0.1995</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9520</td>
<td>0.9630</td>
<td>0.9570</td>
<td>0.9530</td>
<td>0.9490</td>
<td>0.9480</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.0059</td>
<td>0.0036</td>
<td>-0.0048</td>
<td>0.0148</td>
<td>0.0132</td>
<td>0.0174</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3405</td>
<td>0.3374</td>
<td>0.2943</td>
<td>0.3659</td>
<td>0.3651</td>
<td>0.3481</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.3522</td>
<td>0.3495</td>
<td>0.3028</td>
<td>0.3537</td>
<td>0.3524</td>
<td>0.3371</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9660</td>
<td>0.9680</td>
<td>0.9650</td>
<td>0.9500</td>
<td>0.9510</td>
<td>0.9500</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>0.0120</td>
<td>0.0101</td>
<td>0.0080</td>
<td>-0.0002</td>
<td>-0.0003</td>
<td>0.0006</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1710</td>
<td>0.1500</td>
<td>0.1359</td>
<td>0.1624</td>
<td>0.1441</td>
<td>0.1406</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1677</td>
<td>0.1456</td>
<td>0.1328</td>
<td>0.1650</td>
<td>0.1454</td>
<td>0.1404</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9510</td>
<td>0.9510</td>
<td>0.9500</td>
<td>0.9620</td>
<td>0.9600</td>
<td>0.9480</td>
</tr>
<tr>
<td>0.5</td>
<td>500</td>
<td>-0.0058</td>
<td>-0.0057</td>
<td>-0.0028</td>
<td>0.0081</td>
<td>0.0072</td>
<td>0.0052</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2455</td>
<td>0.2443</td>
<td>0.2088</td>
<td>0.2424</td>
<td>0.2416</td>
<td>0.2304</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2468</td>
<td>0.2452</td>
<td>0.2130</td>
<td>0.2470</td>
<td>0.2464</td>
<td>0.2358</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9470</td>
<td>0.9460</td>
<td>0.9520</td>
<td>0.9540</td>
<td>0.9530</td>
<td>0.9590</td>
</tr>
<tr>
<td></td>
<td>1000</td>
<td>-0.0045</td>
<td>-0.0047</td>
<td>-0.0054</td>
<td>0.0060</td>
<td>0.0047</td>
<td>0.0023</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1770</td>
<td>0.1545</td>
<td>0.1347</td>
<td>0.1785</td>
<td>0.1551</td>
<td>0.1496</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.1783</td>
<td>0.1552</td>
<td>0.1402</td>
<td>0.1757</td>
<td>0.1551</td>
<td>0.1492</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9490</td>
<td>0.9530</td>
<td>0.9610</td>
<td>0.9420</td>
<td>0.9470</td>
<td>0.9450</td>
</tr>
<tr>
<td>0.5</td>
<td>500</td>
<td>0.0113</td>
<td>0.0094</td>
<td>0.0092</td>
<td>0.0034</td>
<td>0.0024</td>
<td>0.0065</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2752</td>
<td>0.2743</td>
<td>0.2321</td>
<td>0.2734</td>
<td>0.2728</td>
<td>0.2616</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.2713</td>
<td>0.2687</td>
<td>0.2329</td>
<td>0.2725</td>
<td>0.2715</td>
<td>0.2596</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.9470</td>
<td>0.9480</td>
<td>0.9610</td>
<td>0.9480</td>
<td>0.9460</td>
<td>0.9430</td>
</tr>
</tbody>
</table>

Note: $mr = 40\%$. The true parameters $\beta = (0.5, \log(2))^T$.
Table 2
Simulation results of Case 2 (n = 1000).

<table>
<thead>
<tr>
<th>p</th>
<th>$\beta_0$ Bias</th>
<th>$\beta_{wp}$</th>
<th>$\beta_{ws}$</th>
<th>$\beta_1$ Bias</th>
<th>$\beta_{wp}$</th>
<th>$\beta_{ws}$</th>
<th>$\beta_2$ Bias</th>
<th>$\beta_{wp}$</th>
<th>$\beta_{ws}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0090</td>
<td>0.0083</td>
<td>0.0092</td>
<td>-0.0031</td>
<td>-0.0038</td>
<td>-0.0023</td>
<td>-0.0106</td>
<td>-0.0108</td>
<td>-0.0129</td>
</tr>
<tr>
<td></td>
<td>0.1035</td>
<td>0.0998</td>
<td>0.0939</td>
<td>0.1327</td>
<td>0.1155</td>
<td>0.0945</td>
<td>0.1413</td>
<td>0.1399</td>
<td>0.1298</td>
</tr>
<tr>
<td></td>
<td>0.1062</td>
<td>0.1017</td>
<td>0.0964</td>
<td>0.1359</td>
<td>0.1180</td>
<td>0.0976</td>
<td>0.1451</td>
<td>0.1439</td>
<td>0.1358</td>
</tr>
<tr>
<td></td>
<td>0.9530</td>
<td>0.9520</td>
<td>0.9520</td>
<td>0.9600</td>
<td>0.9550</td>
<td>0.9500</td>
<td>0.9510</td>
<td>0.9570</td>
<td>0.9530</td>
</tr>
<tr>
<td>0.7</td>
<td>0.0125</td>
<td>0.0087</td>
<td>0.0060</td>
<td>0.0115</td>
<td>0.0085</td>
<td>0.0023</td>
<td>-0.0153</td>
<td>-0.0204</td>
<td>-0.0153</td>
</tr>
<tr>
<td></td>
<td>0.1509</td>
<td>0.1432</td>
<td>0.1340</td>
<td>0.1958</td>
<td>0.1704</td>
<td>0.1401</td>
<td>0.2167</td>
<td>0.2142</td>
<td>0.1951</td>
</tr>
<tr>
<td></td>
<td>0.1512</td>
<td>0.1450</td>
<td>0.1362</td>
<td>0.1909</td>
<td>0.1677</td>
<td>0.1369</td>
<td>0.2107</td>
<td>0.2107</td>
<td>0.1939</td>
</tr>
<tr>
<td></td>
<td>0.9550</td>
<td>0.9570</td>
<td>0.9500</td>
<td>0.9490</td>
<td>0.9440</td>
<td>0.9400</td>
<td>0.9410</td>
<td>0.9460</td>
<td>0.9540</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0093</td>
<td>0.0094</td>
<td>0.0106</td>
<td>-0.0030</td>
<td>0.0049</td>
<td>0.0018</td>
<td>-0.0243</td>
<td>-0.0247</td>
<td>-0.0268</td>
</tr>
<tr>
<td></td>
<td>0.2015</td>
<td>0.1944</td>
<td>0.1760</td>
<td>0.2476</td>
<td>0.2152</td>
<td>0.1788</td>
<td>0.3057</td>
<td>0.3002</td>
<td>0.2748</td>
</tr>
<tr>
<td></td>
<td>0.2103</td>
<td>0.2017</td>
<td>0.1826</td>
<td>0.2635</td>
<td>0.2318</td>
<td>0.1885</td>
<td>0.2970</td>
<td>0.2902</td>
<td>0.2748</td>
</tr>
<tr>
<td></td>
<td>0.9630</td>
<td>0.9580</td>
<td>0.9600</td>
<td>0.9750</td>
<td>0.9740</td>
<td>0.9650</td>
<td>0.9370</td>
<td>0.9360</td>
<td>0.9470</td>
</tr>
<tr>
<td>0.7</td>
<td>0.0243</td>
<td>0.0247</td>
<td>0.0268</td>
<td>0.0294</td>
<td>0.0276</td>
<td>0.0114</td>
<td>-0.0304</td>
<td>-0.0298</td>
<td>-0.0340</td>
</tr>
<tr>
<td></td>
<td>0.3199</td>
<td>0.3112</td>
<td>0.2837</td>
<td>0.3496</td>
<td>0.3451</td>
<td>0.2983</td>
<td>0.3570</td>
<td>0.3502</td>
<td>0.3503</td>
</tr>
<tr>
<td></td>
<td>0.3057</td>
<td>0.3002</td>
<td>0.2748</td>
<td>0.3503</td>
<td>0.3473</td>
<td>0.2928</td>
<td>0.2970</td>
<td>0.2947</td>
<td>0.2961</td>
</tr>
<tr>
<td></td>
<td>0.9370</td>
<td>0.9360</td>
<td>0.9470</td>
<td>0.9610</td>
<td>0.9600</td>
<td>0.9580</td>
<td>0.2308</td>
<td>0.2341</td>
<td>0.2308</td>
</tr>
<tr>
<td></td>
<td>0.3308</td>
<td>0.3318</td>
<td>0.2775</td>
<td>0.3626</td>
<td>0.3600</td>
<td>0.2806</td>
<td>0.3270</td>
<td>0.3241</td>
<td>0.3229</td>
</tr>
<tr>
<td></td>
<td>0.9600</td>
<td>0.9670</td>
<td>0.9560</td>
<td>0.9650</td>
<td>0.9630</td>
<td>0.9510</td>
<td>0.9600</td>
<td>0.9670</td>
<td>0.9560</td>
</tr>
</tbody>
</table>

Note: corr(X, W) = 0.79. The true parameters $\beta = (0.5, 1, -1)^T$.

indicator $d$ was a binary variable with $P(d = 1|Y^0, W, Z) = H(x_0 + x_1) + x_2W + x_3Z)$, where the values of $x$ were set at $(0.5, -0.5, -0.5, 0.5)$ and $(1, 1, 0.5, -1)$. This setup resulted in about 24% and 45% of missing rate, respectively. The corr(X, W) is about 0.79.

Simulation results (see Table 2) show the efficiencies of all the estimators increase as $p$ increases. When $p = 1$, all the estimators have smaller standard deviation compared to the case of $p = 0.7$ and 0.5. Note that when $p = 1$, all the data are from direct question (i.e., $Y^0 = Y$) the case considered by Wang and Wang (1997). Table 2 shows that the parameter estimates from direct question generally have smaller standard errors than the randomized response counterparts. For the estimation of $\beta_1$, all estimates perform better when $mr = 24\%$ compared to $mr = 45\%$. Overall, the efficiencies of all the estimators increase as the missing rate decreases. In general, the $\hat{\beta}_{ws}$ is more efficient than the other two estimators $\hat{\beta}_{wp}$ and $\hat{\beta}_t$.

In summary, we conclude that $\hat{\beta}_{ws}$ is asymptotically more efficient than $\hat{\beta}_{wp}$ and $\hat{\beta}_t$. Moreover, the $\hat{\beta}_{wp}$ is asymptotically more efficient than $\hat{\beta}_t$. The efficiencies of all the estimators are impacted by the values of $p$, $c$, $n$, and missing rate. However, if the missing rate is high and randomizing proportion $p$ is low, we are not able to obtain enough information about the question of interest, which can result in divergence of the estimator. The other problem of the weighted estimator is the danger of having very small estimated selection probabilities which as a result give very large weight. To deal with this
issue, it is customary to collapse classes by certain criteria (e.g. Thomsen, 1973; Little, 1986; Eltinge and Yansanch, 1997; Haziza and Beaumont, 2007).

6. Example

The data are from a customer survey on cable TV subscribers in Taiwan. To gain some knowledge about the proportion of illegal cable TV subscribers, the unrelated question RRT design was used. First, respondents were asked to generate a random number between 31 and 89. Then, the following two questions were asked:

Q1: Is the generate random number above 70?
Q2: Is your cable TV illegally connected?

If the number is an odd number, please answer Q1; otherwise answer Q2. The probability of answering Q2 is equal to \( \frac{29}{40} \) (i.e. 0.725). The probability of answering “Yes” of Q2 is equal to \( \frac{10}{40} \) (i.e. 0.25). Let \( Y^0 \) be the response of one question Q1 or Q2. The covariate \( X \) is the response of the question “Are you satisfied with the cable TV maintenance?”. In this study, due to item nonresponse, \( X \) is not available for some subjects. A surrogate variable for \( X \), denoted by \( W \), is the three cities in Taiwan (1. Tainan; 2. Kaohsiung; 3. Pingtung), where \( DW_j \) (\( j = 1, 2 \)) is the dummy variable for \( W \), where \( DW_j = 1 \) if \( W = j \) and zero otherwise. The covariate \( Z \) is the response of the question “Are you satisfied with the cable TV channel?”. A random sample of 3159 subjects are selected from three cities, and there are 2465 subjects in the validation data set.

To verify if the missing mechanism is MAR, we conduct a logistic regression analysis with outcome variables \( \delta_i \) (for \( i = 1, 2, \ldots, n \) and \( n = 3159 \)) and four covariates \( (Y^0, DW_1, DW_2, Z) \). The parameter estimates for these four covariates are equal to (0.788, 0.719, 0.881, 0.765), with corresponding standard error (SE) (0.099, 0.111, 0.076, 0.074). Hence the missingness mechanism depends on the joint distributions of \( Y^0, W, \) and \( Z \) (i.e. MAR).

We examine the performance of the weighted semiparametric estimator \( \hat{\beta}_{ws} \) and weighted parametric estimator \( \hat{\beta}_{wp} \). Table 3 shows the estimates of \( (\beta_k, k = 0, 1, 2) \) and the corresponding estimated standard errors (SE) of the estimators. The estimated SE of \( \hat{\beta}_{ws} \) is smaller than that of the \( \hat{\beta}_{wp} \), which is consistent with simulation results in Section 5.

<table>
<thead>
<tr>
<th>( \beta_k )</th>
<th>( \hat{\beta}_{ws} )</th>
<th>( \hat{\beta}_{wp} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_0 )</td>
<td>-2.025</td>
<td>-2.167</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.453)</td>
<td>(0.494)</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>-0.978</td>
<td>-1.023</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.467)</td>
<td>(0.502)</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>-0.827</td>
<td>-0.771</td>
</tr>
<tr>
<td>(SE)</td>
<td>(0.351)</td>
<td>(0.366)</td>
</tr>
</tbody>
</table>

7. Conclusions

We have proposed the Horvitz and Thompson (1952)-type weighted estimators for logistic regression analysis of randomized response data where some covariates are missing at random. Specifically, we compare the efficiencies of several estimators of the logistic regression by using different estimators of the selection probabilities. Both theoretical analysis and simulation results conclude that the proposed estimators \( \hat{\beta}_{wp} \) and \( \hat{\beta}_{wp} \) are more efficient than the \( \hat{\beta}_{ws} \). This phenomenon is quite useful, namely, plugging in estimated selection probabilities is better than using the true values of selection probabilities even if we know the true values. Furthermore, simulation results conclude that the proposed estimator \( \hat{\beta}_{ws} \) is more efficient than the \( \hat{\beta}_{wp} \) estimators.

An important feature of our proposed method is that no assumptions are made about the distribution of the selection probability. Our simulation results also conclude that the parameter estimates from direct question data (i.e. \( p = 1 \)) generally have smaller standard errors than the randomized response counterparts. Although the main results were presented for the case when both \( W \) and \( Z \) are discrete. Extension to continuous cases can be done by the approach of Wang and Wang (1997). In the future, it will be interesting to extend the weighted method to the case of different missing mechanisms, e.g. \( Y^0 \) is missing or both \( W \) and \( Y^0 \) are missing.

Moreover, when \( Y^0 = Y \), as discussed by Wang et al. (2007), it can be shown that when the observed \( X \) and \( Y \) are categorical, the simple weighted estimator, augmented inverse probability weighted estimator (Robins et al., 1994) and mean score estimator (Reilly and Pepe, 1995) are equivalent to one another. Similarly, under unrelated-question RRT, when \( X \) is MAR, \( V \) is categorical and \( \pi(Y^0, V) \) is used in (7), it is easy to show that both augmented inverse probability weighted estimator and mean score estimator reduce to the weighted estimator \( \hat{\beta}_{ws} \).
Appendix A

We assume that the regularity Conditions (A1)-(A5) of Section 4 hold. In this appendix we prove Lemma 1, Theorems 1 and 2. Note that Condition (A1) is a common assumption for the existence of the inverse of the inclusion selection probabilities. Condition (A2) is a usual assumption for the unique solution of estimating equation. Condition (A3) is a usual assumption for the existence of variance. Condition (A4) is a usual assumption for the proof of consistency in estimating equation theory. Condition (A5) is a usual assumption for uniqueness of \( \alpha \) solution of estimating equation.

**Proof of Lemma 1.** By Taylor expansion of \( U_n(\hat{\beta}_i, \pi) \) at \( \beta \), we have

\[
0 = U_n(\hat{\beta}_i, \pi) = U_n(\beta, \pi) + \frac{\partial U_n(\beta, \pi)}{\partial \beta^T}(\hat{\beta}_i - \beta) + o_p(1).
\]

Let \( G_n(\beta, \pi) = (1/n)[-\partial U_n(\beta, \pi)/\partial \beta] \). Then

\[
\sqrt{n}(\hat{\beta}_i - \beta) = G_n(\beta, \pi)^{-1}U_n(\beta, \pi) + o_p(1).
\]

Since the estimating equation (4) is unbiased, this ensures that \( \hat{\beta}_i \) is a consistent estimator of \( \beta \). Since \( G_n(\beta, \pi) \Rightarrow G(\beta, \pi) \) under Condition (A2), \( G(\beta, \pi) \) is nonsingular, we have \( \sqrt{n}(\hat{\beta}_i - \beta) = G(\beta, \pi)^{-1}U_n(\beta, \pi) + o_p(1) \). By Condition (A3), it can be shown that \( \text{Cov}(U_n(\beta, \pi)) = 0 \). Using Central Limit Theorem, \( \sqrt{n}(\hat{\beta}_i - \beta) \) is asymptotically normally distributed with mean 0 and covariance matrix

\[
\Delta_n = G^{-1}(\beta, \pi)J(\beta, \pi)G^{-T}(\beta, \pi).
\]

**Proof of Theorem 1.** We first prove that \( \hat{\beta}_{wp} \) is a consistent estimator of \( \beta \). Now, let

\[
G_n(\beta, \pi) = \frac{1}{n} \left[ -\frac{\partial U_n(\beta, \pi)}{\partial \beta^T} \right]
\]

and

\[
D_n(\beta, \pi) = \frac{1}{n} \left[ -\frac{\partial U_n(\beta, \pi)}{\partial \pi} \right]
\]

where

\[
B_i(\beta) = \frac{PH^{(1)}(\beta, \pi_i)\pi_i}{PH(\beta^T \pi_i)} \left\{ \frac{[1 - 2PH(\beta^T \pi_i)] - PH^{(1)}(\beta, \pi_i) \pi_i}{[1 - PH(\beta^T \pi_i)]^2} \right\}.
\]

Then by law of large number, it can be shown that \( G_n(\beta, \pi) \Rightarrow G(\beta, \pi) \) and \( D_n(\beta, \pi) \Rightarrow D(\beta, \pi) \). By Condition (A4), the convergence of \( G_n(\beta, \pi) \) to \( G(\beta, \pi) \) is uniform in a neighborhood of the true \( \beta \). By the Inverse Function Theorem of Foutz (1977), along with Condition (A2), there exists a unique consistent solution to the estimating equation \( U_n(\beta, \pi) = 0 \) in a neighborhood of the true \( \beta \). Hence it follows that \( \hat{\beta}_{wp} \) is a consistent estimator of \( \beta \).

Next, we derive the asymptotic distribution of \( \sqrt{n}(\hat{\beta}_{wp} - \beta) \). By Taylor series expansion of \( U_n(\hat{\beta}_{wp}, \hat{\alpha}) \) at \( (\beta, \pi) \), we have

\[
0 = U_n(\hat{\beta}_{wp}, \hat{\alpha}) = U_n(\beta, \pi) + \frac{\partial U_n(\beta, \pi)}{\partial \beta^T}(\hat{\beta}_{wp} - \beta) + \frac{\partial U_n(\beta, \pi)}{\partial \pi}(\hat{\alpha} - \alpha) + o_p(1)
\]

\[
= U_n(\beta, \pi) - G_n(\beta, \pi)\sqrt{n}(\hat{\beta}_{wp} - \beta) - D_n(\beta, \pi)\sqrt{n}(\hat{\alpha} - \alpha) + o_p(1).
\]

In addition, under model (5), the MLE of \( \pi \) is the solution of \( M_n(\pi) = 1/\sqrt{n} \sum_{i=1}^{n} \pi_{i}(\hat{\alpha}_i - \pi_i(\alpha)) = 0 \). Since the estimating equation is unbiased, the \( \hat{\alpha} \) is a consistent estimator of \( \alpha \). By Taylor series expansion \( \sqrt{n}(\hat{\alpha} - \alpha) = K^{-1}(\alpha)M_n(\pi) + o_p(1) \), where

\[
M_n(\pi) = 1 - \frac{1}{\sqrt{n}}[\pi - \pi(\alpha)] = n^{-1} \sum_{i=1}^{n} \pi_i(\alpha)(\hat{\alpha}_i - \pi_i(\alpha)).
\]

Using law of large number and Central Limit Theorem, we have \( K^{-1}(\alpha) \Rightarrow K(\alpha) \) and \( M_n(\pi) \) converges in distribution to \( \mathcal{N}(0, K(\alpha)) \), where \( K(\alpha) \) was defined in Theorem 1. By Condition (A5), \( \sqrt{n}(\hat{\alpha} - \alpha) \) is asymptotically normally distributed with mean 0 and variance \( K^{-1}(\alpha) \). Since \( G_n(\beta, \pi) \Rightarrow G(\beta, \pi) \) and \( D_n(\beta, \pi) \Rightarrow D(\beta, \pi) \), we have

\[
\sqrt{n}(\hat{\beta}_{wp} - \beta) = G^{-1}(\beta, \pi)U_n(\beta, \pi) - D(\beta, \pi)K^{-1}(\alpha)M_n(\pi) + o_p(1).
\]

By Condition (A3), it can be shown that \( \text{Cov}(U_n(\beta, \pi)) = 0 \) and \( \text{Cov}(U_n(\beta, \pi), M_n(\pi)) = D(\beta, \pi) \). Using Central Limit Theorem, \( \sqrt{n}(\hat{\beta}_{wp} - \beta) \) is asymptotically normally distributed with mean 0 and covariance matrix

\[
\Delta_{wp} = G^{-1}(\beta, \pi)[D(\beta, \pi)K^{-1}(\alpha)D^T(\beta, \pi)]G^{-T}(\beta, \pi).
\]
Proof of Theorem 2. Define \( S(\beta) = \mathcal{X}_i \mathcal{A}_i(\beta)[Y_i^0 - pH(V, Y_i)] - k \). Then

\[
U_n(\beta, \hat{\beta}) - U_n(\beta, \pi) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{1}{\pi(Y_i^0, V_i)} - \frac{1}{\pi(Y_i^0, V_i)} \right\} \delta S_i(\beta)
\]

\[
= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \hat{\pi}(Y_i^0, V_i) - \pi(Y_i^0, V_i) \right\} \delta S_i(\beta) + O_p(1) \frac{1}{\sqrt{n}}
\]

\[
= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \hat{\pi}(Y_i^0, V_i) - \pi(Y_i^0, V_i) \right\} \delta S_i(\beta) + O_p(1) \frac{1}{\sqrt{n}}
\]

\[
= -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{1}{\pi(Y_i^0, V_i)} \delta_i(Y_i^0, V_i) \right\} \delta S_i(\beta) + O_p(1) \frac{1}{\sqrt{n}}
\]

Let \( S(\beta) = E[S_i(\beta)|Y_i^0 = Y_i^0, V_i = V_i] \). It is easily seen that

\[
D_{2n} = \frac{1}{n^{3/2}} \sum_{j=1}^{n} \frac{\pi(Y_j^0, V_j) \delta_j(Y_j^0, V_j) |\delta_j|}{\pi^2(Y_j^0, V_j) \delta_j Y_j^0, V_j = V_j} S_i(\beta) + O_p(1) \frac{1}{\sqrt{n}}
\]

\[
= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\delta_j(Y_j^0, V_j)}{\pi(Y_j^0, V_j)} \frac{1}{\pi(Y_j^0, V_j)} \delta_j Y_j^0, V_j = V_j S_i(\beta)
\]

\[
= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \frac{\delta_j(Y_j^0, V_j)}{\pi(Y_j^0, V_j)} \left[ S_i(\beta) + O_p(1) \frac{1}{\sqrt{n}} \right]
\]

In addition, let

\[
l_{ij} = \frac{\delta_i(Y_i^0, V_i) |\delta_i| \delta_j(Y_j^0, V_j) |\delta_j|}{\pi^2(Y_i^0, V_i) \delta_i Y_i^0, V_i = V_i} S_i(\beta)
\]

In order to show \( D_{1n} = O_p(1/\sqrt{n}) \). We will show \( E(D_{1n}) = O(1/\sqrt{n}) \) and \( \text{Var}(D_{1n}) = O(1/n) \). First, note that

\[
E \left\{ \left[ \frac{\delta_i(Y_i^0, V_i) |\delta_i| \delta_j(Y_j^0, V_j) |\delta_j|}{\pi^2(Y_i^0, V_i) \delta_i Y_i^0, V_i = V_i} S_i(\beta) \right] |X_i, Y_j^0 = Y_i^0, V_j = V_j \right\}
\]

\[
= \begin{cases} 
0 & \text{if } i \neq j, \\
E \left\{ \frac{\pi(Y_j^0, V_j) (1 - \pi(Y_j^0, V_j)) S_i(\beta) |\delta_j|}{\pi^2(Y_j^0, V_j) \delta_j Y_j^0, V_j = V_j} \right\} & \text{if } i = j,
\end{cases}
\]
Hence, we have

\[
E(D_{1n}) = \frac{1}{n^{3/2}} \sum_{i=1}^{n} \sum_{j=1}^{n} \left\{ \frac{\left( \delta_i - \pi(Y_i^0, V_i) \right) \delta_j - \pi(Y_j^0, V_j) \mid \pi(Y_j^0, V_j) \mid \pi(Y_j^0, V_j) \mid Y_j^0 = Y_i^0, V_j = V_i \} \right. \}
\]

By direct calculation we have

\[
\text{Cov}(l_{ij}, l_{kl}) = \left\{ \frac{\left( \delta_i - \pi(Y_i^0, V_i) \right) \delta_j - \pi(Y_j^0, V_j) \mid \pi(Y_j^0, V_j) \mid \pi(Y_j^0, V_j) \mid Y_j^0 = Y_i^0, V_j = V_i \} \right. \}
\]

Hence,

\[
\text{Var}(D_{1n}) = \frac{1}{n^3} \left\{ \sum_{ij} n \text{Var}(l_{ij}) + \sum_{ij} n \text{Cov}(l_{ij}, l_{kl}) + \sum_{k=1, k \neq j}^{n} \text{Cov}(l_{ij}, l_{kl}) + \sum_{k=1, k \neq j}^{n} \text{Cov}(l_{ij}, l_{kl}) \right\}
\]

Therefore, \( D_{1n} = O_p(1/\sqrt{n}) \) and

\[
U_n(\beta, \pi) - U_n(\hat{\beta}, \pi) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{\delta_i - \pi(Y_i^0, V_i) \mid \pi(Y_i^0, V_i) \mid Y_i^0, V_i \} \right. \}
\]

we have

\[
U_n(\beta, \pi) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left\{ \frac{\delta_i S_i(\beta)}{\pi(Y_i^0, V_i)} - \frac{\delta_i - \pi(Y_i^0, V_i) \mid \pi(Y_i^0, V_i) \mid Y_i^0, V_i \} \right. \}
\]

Then it can be shown that \( G_n(\beta, \pi) \rightarrow G(\beta, \pi) \), where \( G(\beta, \pi) \) was defined in Theorem 2. By Condition (A4), the convergence of \( G_n(\beta, \pi) \) to \( G(\beta, \pi) \) is uniform in a neighborhood of the true \( \beta \). By the Inverse Function Theorem of Foutz (1977) and Condition (A2), there exists an unique consistent solution to the estimating equation \( U_n(\beta, \pi) = 0 \) in a neighborhood of the true \( \beta \). Hence it follows that \( \beta_{\text{mle}} \) is a consistent estimator of the \( \beta \).
Next, we derive the asymptotic distribution of $\sqrt{n}(\hat{\beta}_{ws} - \beta)$. By Taylor expansion of $U_n(\hat{\beta}_{ws}, \hat{\pi})$ at $(\beta, \hat{\pi})$, we have

$$0 = U_n(\hat{\beta}_{ws}, \hat{\pi})$$

$$= U_n(\beta, \hat{\pi}) + \frac{\partial U_n(\beta, \hat{\pi})}{\partial \beta} (\hat{\beta}_{ws} - \beta) + o_p(1).$$

Since $C_n(\beta, \hat{\pi})^T \Gamma_n C_n(\beta, \hat{\pi})$ under Condition (A2), we have

$$\sqrt{n}(\hat{\beta}_{ws} - \beta) = \Gamma_n^{-1}(U_n(\beta, \hat{\pi}) + [U_n(\beta, \hat{\pi}) - U_n(\beta, \pi)]) + o_p(1).$$

First, we rewrite the $\sqrt{n}(\hat{\beta}_{ws} - \beta) = \Gamma_n^{-1}(U_n(\beta, \hat{\pi}) + [U_n(\beta, \hat{\pi}) - U_n(\beta, \pi)]) + o_p(1)$. Next, we calculate the variance of $U_n(\beta, \hat{\pi})$. By Condition (A3), it is easy to see that

$$\text{Cov}[U_n(\beta, \pi)] = E \left\{ \frac{S_1(\beta) S_1^T(\beta)}{\pi(Y_1^0, V_1)} \right\}.$$ 

Also,

$$\text{Cov}[U_n(\beta, \hat{\pi}), U_n(\beta, \hat{\pi}) - U_n(\beta, \pi)] = -E \left\{ \frac{d_1[ 1 - \pi(Y_0^0, V_1)] S_1(\beta) S_1^T(\beta)}{\pi^2(Y_1^0, V_1)} \right\} + o(1)$$

$$= -E \left\{ \frac{1 - \pi(Y_0^0, V_1)}{\pi(Y_1^0, V_1)} S_1(\beta) S_1^T(\beta) \right\} + o(1).$$

Moreover,

$$\text{Cov}[U_n(\beta, \hat{\pi}) - U_n(\beta, \pi)] = E \left\{ \frac{[1 - \pi(Y_0^0, V_1)]^2 S_1(\beta) S_1^T(\beta)}{\pi^2(Y_1^0, V_1)} \right\} + o(1)$$

$$= E \left\{ \frac{1 - \pi(Y_0^0, V_1)}{\pi(Y_1^0, V_1)} S_1(\beta) S_1^T(\beta) \right\} + o(1).$$

Therefore, we have

$$\text{Cov}[U_n(\beta, \hat{\pi})] = E \left\{ \frac{S_1(\beta) S_1^T(\beta)}{\pi(Y_1^0, V_1)} \right\} - E \left\{ \frac{1 - \pi(Y_0^0, V_1)}{\pi(Y_1^0, V_1)} S_1(\beta) S_1^T(\beta) \right\} + o(1).$$

Let $J(\beta, \pi) = E[S_1(\beta) S_1^T(\beta)/\pi(Y_1^0, V_1)]$, $J^*(\beta, \pi) = E[S_1(\beta) S_1^T(\beta)/\pi(Y_0^0, V_1)]$, and $C^*(\beta, \pi) = E[S_1(\beta) S_1^T(\beta)]$. Then we can obtain

$$A_{ws} = G^{-1}(\beta, \pi)(J(\beta, \pi) - J^*(\beta, \pi) - C^*(\beta, \pi))G^{-T}(\beta, \pi).$$

References


