Statistica Sinica **35** (2025), 25-48
doi:https://doi.org/10.5705/ss.202022.0320

REGRESSION ANALYSIS OF RANDOMIZED RESPONSE EVENT TIME DATA

Chi-Chung Wen* and Yi-Hau Chen

Tamkang University and Academia Sinica

Abstract: The randomized response technique (RRT) is used to reduce underreporting of sensitive characteristics in survey studies by enhancing privacy protection. Currently, the RRT is mainly applied for prevalence estimation of some sensitive event. We extend the application of the RRT to an analysis of a time-to-event outcome. Event time data collected from surveys are usually subject to case-I interval-censoring, so that only "current-status" data on the occurrence of the event by the examination time are available. As such, we focus on current-status (case-I interval-censored) event time data collected using the RRT. Based on the data, we propose a semiparametric maximum likelihood estimation procedure for the event time distribution given the covariates. The proposed method is assumed to follow a general class of semiparametric transformation models characterized by a parametric function for the relationship between the event time and the covariates, as well as an unspecified baseline function. We develop the asymptotic theory for the proposed estimation, including the consistency and asymptotic normality, and examine its finite-sample properties using simulation studies. We apply the proposed method to current-status data surveyed using the RRT to make statistical inferences on the time to incidence of extramarital relations since marriage.

Key words and phrases: Randomized response technique, semiparametric maximum likelihood estimation, semiparametric transformation model, sensitive issue.

1. Introduction

The time duration to the occurrence of some event of interest is a common and important research target in various disciplines, including medicine and the social sciences. Usually, the observation of the event time is subject to incompleteness caused by a limitation of the observation procedure, a well-known phenomenon called "censoring." For example, when the occurrence of an event is not tracked continuously, but only by a sequence of examinations, the event time lies in some interval between two consecutive examinations, that is, the event time is subject to interval censoring. In survey studies, it is common for there to be only one examination during the observation period. Here, the event time is subject to a special case of interval censoring, called case-I interval censoring, and the resulting event time data are called "current-status" data. Analyses of case-I interval-censoring data, including regression analyses, are widely used in

*Corresponding author. E-mail: ccwen@mail.tku.edu.tw

the literature; see Huang (1996), Lin, Oakes and Ying (1998), Martinussen and Scheike (2002), Sun and Sun (2005), Tian and Cai (2006), Sun (2006), Zhang and Sun (2010), Wen and Chen (2012), among others, for further details.

In addition to censoring, event time data obtained from survey studies are also subject to participants' response bias, that is, the participants answer the survey questions inaccurately or falsely. Response bias is of particular concern when the survey questions are related to moral, legal, or other sensitive issues, because there is a greater chance that participants will want to protect their privacy, and thus conceal the truth on such issues. In other words, response data from sensitive survey questions can be highly erroneous.

To encourage respondents to truthfully answer sensitive questions, the randomized response technique (RRT) was proposed by Warner (1965), and has been extended by various subsequent authors (Greenberg et al., 1969; Horvitz, Shah and Simmons, 1967; Kuk, 1990; Singh, Singh and Mangat, 2000; Gjestvang and Singh, 2006; Narjis and Shabbir, 2023). The RRT encourages respondents to respond to sensitive issues by using some random device (such as a coin, cards, or dice), the outcome of which is blind to the interviewer. As a result, the respondents may feel it is safer to tell the truth, because they cannot be identified by the interviewer, thus reducing underreporting of sensitive characteristics in surveys (Scheers and Dayton, 1988).

Two types of the RRT are popular: the related- and unrelated-question RRTs. Warner's (1965) original RRT proposal is now termed the "related-question" RRT, and uses a question set consisting of a sensitive question A, and a complementary question A^c. For example,

A: *Have you ever had sex with someone other than your spouse?*
A^c: *Have you never had sex with someone other than your spouse?*

The interviewee is asked either A or A^c, determined by a random device and unknown to the interviewer. Greenberg et al. (1969) proposed the unrelated-question RRT, which employs a question set consisting of a sensitive question A, and an innocuous question B. For example,

A: *Have you ever had sex with someone other than your spouse?*
B: *Were you born in the month of January, February, or March?*

The interviewee answers one of A or B, which again is determined by a random device and unknown to the interviewer. The probability of answering "yes" to question B is known, or can be estimated before the survey. A simple RRT is designed by asking interviewees to either answer question A truthfully, or answer "yes," regardless of the truth, according to the outcome of a random device. This RRT can also be conceived as an unrelated-question RRT with an innocuous question B, such that the probability of answering "yes" to B is one.

Although the RRT has been applied to obtain less biased estimations of the prevalence (Warner, 1965; Greenberg et al., 1969) and factors (Scheers and Dayton, 1988) affecting sensitive behavior, it has not been applied to estimate an event time distribution. As mentioned above, event times in survey studies are usually only observed in the form of interval-censored data, often case-I interval-censored, or current-status data. Hence, an analysis of time-to-event data collected using the RRT may also need to address the interval-censoring problem in addition to the special data structure under the RRT.

In this work, we propose a general methodology for analyzing current-status event time data collected using the RRT in a survey study. Specifically, we propose a semiparametric maximum likelihood estimation procedure for the event time distribution given the covariates, which is assumed to follow semiparametric transformation models characterized by a parametric function for the relationship between the event time and the covariates, and an unspecified baseline function. We also discuss the asymptotic theory, including the consistency and asymptotic normality, for the proposed estimation, and examine its finite-sample properties using simulation studies. We apply the proposed method to a set of current-status data collected using the RRT to make statistical inferences about the time to the incidence of extramarital relations after marriage.

2. Current-Status Data and Model

Let T, C, and Z denote the time to the sensitive event of interest, survey time, and covariate vector for an individual, respectively. Then, $\delta \equiv I(T \leq C)$ indicates whether or not the sensitive event has occurred by the survey time C, where $I(\cdot)$ is the indicator function.

Given the covariate Z, the event time T is assumed to follow a semiparametric transformation model, namely, at time t, the conditional distribution function of T given Z is of the form

$$\Pr(T \leq t|Z = z) = F(\exp(\beta' z)H(t)), \tag{2.1}$$

where F is a known distribution function, β is an unknown vector of regression parameters, and H is an unspecified increasing real-valued function. The choices of $F(x) = 1-\exp(-x)$ and $F(x) = x/(1+x)$ lead to the proportional hazards and proportional odds model, respectively, which are well-known models for lifetime distributions with ranges on [0,1]; see, for example, Zeng and Lin (2006) and Zeng and Lin (2007).

Throughout the paper, we assume the censoring mechanism is noninformative, that is, T and C are conditionally independent given Z.

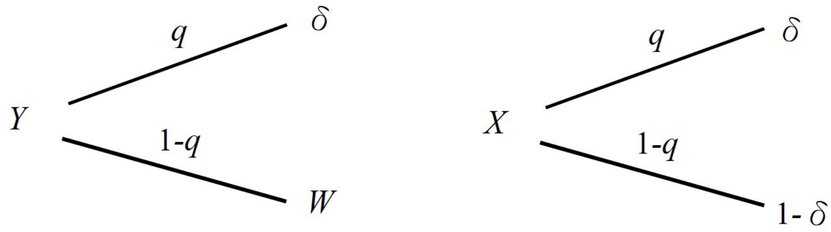

Figure 1. Diagrams of the probability mechanism for the unrelated-question RRT (left panel) and the related-question RRT (right panel).

3. Estimation under the Unrelated RRT

Suppose that the current-status data for the sensitive event are collected using the unrelated-question RRT (Greenberg et al., 1969). Recall that the unrelated-question RRT consists of a question A on the sensitive event and an innocuous question B, and one of them, determined by a random device, is answered by the interviewee. Let Q be a binary random variable, with $Q = 1$ denoting that the interviewee answers question A and $Q = 0$ denoting that the interviewee answers question B, and let W be the binary response of the interviewee to the innocuous question B. A single observation in the unrelated-question RRT survey then consists of $O = \{Y, C, Z\}$, where $Y = Q\delta + (1-Q)W$, with $\delta = I(T \leq C)$. Here, and in the following, we assume that interviewees answer honestly when asked the sensitive question. Figure 1 illustrates the probability mechanism of the answer Y from the unrelated-question RRT. Here, we propose an estimation procedure for the event time model (2.1) based on the independent and identically distributed (i.i.d.) sample $O_i = \{Y_i, C_i, Z_i\}$, for $i = 1, \ldots, n$, of the unrelated-question RRT observation O. Let the probabilities $\Pr(Q = 1) = q$, $\Pr(W = 1|C, Z) = c$, and $\Pr(W = 0|C, Z) = 1 - c$. The values of q and c are assumed known in the following.

Observe that

$$
\begin{aligned}
&\Pr(Y = 1|C, Z)\\
&= \Pr(Y = 1|C, Z, Q = 1)\Pr(Q = 1) + \Pr(Y = 1|C, Z, Q = 0)\Pr(Q = 0)\\
&= \Pr(\delta = 1|C, Z)q + \Pr(W = 1|C, Z)(1 - q)\\
&= qF(e^{\beta' z}H(C)) + c(1 - q).
\end{aligned}
$$

Hence, the likelihood function of $(O_1, \ldots, O_n)$ takes the form

$$L_n(\theta) = \prod_{i=1}^{n} L(\theta|O_i), \tag{3.1}$$

where $L(\theta|O) = \{qF(e^{\beta' z}H(C)) + (1 - q)c\}^Y\{1 - qF(e^{\beta' z}H(C)) - (1 - q)c\}^{1-Y}$

and $\theta = (\beta, H)$. The semiparametric maximum likelihood estimator (SPMLE) $\widehat{\theta} = (\widehat{\beta}, \widehat{H})$, referred to as the RRT estimator, of θ maximizes the likelihood in (3.1).

It is obvious that (3.1) depends on H only through $H(C_i)$, for $i = 1, \ldots, n$. Therefore, in maximizing L_n, we treat $\widehat{H}$ as a right-continuous step function that jumps at the survey time C_i. Let $t_1 < \cdots < t_K$ be the ordered distinct time points of C_i, for $i = 1, \ldots, n$, associated with $Y_i = 1$. For a person with $Y_i = 0$ and $t_k < C_i < t_{k+1}$, consider two increasing step functions, H_1 and H_2, with $H_1 = H_2$ at all t_j, except that $H_1(C_i) = H_1(t_k)$ and $H_2(C_i) > H_1(t_k)$. Because the transformation F is increasing, we conclude that $L(\beta, H_1) > L(\beta, H_2)$. Therefore, $\widehat{H}$ can have jumps only at t_j, for $j = 1, \ldots, K$. Let h_j denote the jump size at t_j.

Because the number K of jumps increases with the sample size, a direct maximization of (3.1) can be challenging. Note that, without randomized response sampling, $O = (Y, C, Z)$ reduces to the current-status data (δ, C, Z). Turnbull (1976) proposed a self-consistency formula for computing the SPMLE, that is essentially an EM algorithm based on current-status data without covariates. Herein, we propose a novel EM algorithm that extends Turnbull's method to a regression analysis of randomized response survival data. Let $N_{ij} = I(T_i \in (t_{j-1}, t_j])$, for $j = 1, \ldots, K$, with $t_0 = 0$. We treat the failure time indicator N_{ij} and the latent indicator Q_i of selecting the sensitive question in the unrelated-question set as missing values in the EM method. Let $\bar{q} = 1 - q$ and $\bar{c} = 1 - c$. The complete-data likelihood of $\{(Y_i, C_i, Z_i, N_{ij}, Q_i), i = 1, \ldots, n, j = 1, \ldots, K\}$ takes the form

$$L_n^c(\theta) = \prod_{i=1}^{n} \Bigg(\Bigg\{ q \prod_{t_j \le C_i} \nabla F(e^{\beta' Z_i} H(t_j))^{N_{ij}} \Bigg\}^{Q_i Y_i} \Big[q \Big\{ 1 - F(e^{\beta' Z_i} H(C_i)) \Big\} \Big]^{Q_i(1-Y_i)} \times (\bar{q}c)^{(1-Q_i)Y_i} (\bar{q}\bar{c})^{(1-Q_i)(1-Y_i)} \Bigg),$$

where $\nabla F(e^{\beta' Z_i} H(t_j)) = F(e^{\beta' Z_i} H(t_j)) - F(e^{\beta' Z_i} H(t_{j-1}))$. In the M-step, we maximize

$$\sum_{i=1}^{n} \Bigg[Y_i \sum_{t_j \le C_i} (N_{ij} Q_i)^{\wedge} \log \nabla F(e^{\beta' Z_i} H(t_j)) + (1 - Y_i) Q_i^{\wedge} \log\{1 - F(e^{\beta' Z_i} H(C_i))\} + Q_i^{\wedge} \log q + (1 - Q_i^{\wedge}) Y_i \log(\bar{q}c) + (1 - Q_i^{\wedge})(1 - Y_i) \log(\bar{q}\bar{c}) \Bigg], \tag{3.2}$$

where

$$Q_i^{\wedge} = E(Q_i | Y_i, C_i, Z_i) = \Pr(Q_i = 1 | Y_i, C_i, Z_i)$$

$$= \frac{qF(e^{\beta' Z_i}H(C_i))^{Y_i}\{1-F(e^{\beta' Z_i}H(C_i))\}^{1-Y_i}}{qF(e^{\beta' Z_i}H(C_i))^{Y_i}\{1-F(e^{\beta' Z_i}H(C_i))\}^{1-Y_i}+\bar{q}c^{Y_i}\bar{c}^{1-Y_i}},$$

and

$$\begin{aligned}(N_{ij}Q_i)^\wedge &= E(N_{ij}Q_i|Y_i,C_i,Z_i) = \Pr(N_{ij}=1,Q_i=1|Y_i,C_i,Z_i)\\ &= \frac{q\nabla F(e^{\beta' Z_i}H(t_j))}{qF(e^{\beta' Z_i}H(C_i))+\bar{q}c}I(t_j\le C_i)Y_i.\end{aligned}$$

To update h_k, for $k = 1\ldots,K$, we use a one-step self-consistency algorithm, as follows. By the first-order approximation, the objective function (3.2) can be approximated by

$$\sum_{i=1}^{n}\Bigg[Y_i\sum_{t_j\le C_i}(N_{ij}Q_i)^\wedge \log\{\dot{F}(e^{\beta' Z_i}H(t_j))e^{\beta' Z_i}h_j\} \tag{3.3}$$

$$+(1-Y_i)Q_i^\wedge \log\{1-F(e^{\beta' Z_i}H(C_i))\}$$

$$+Q_i^\wedge \log q + (1-Q_i^\wedge)Y_i\log(\bar{q}c) + (1-Q_i^\wedge)(1-Y_i)\log(\bar{q}\bar{c})\Bigg], \tag{3.4}$$

where $\dot{F}(x) = dF(x)/dx$. We set the derivative of (3.3) relative to h_k to zero to obtain

$$\begin{aligned}0 = &\sum_{i:Y_i=1}(N_{ik}Q_i)^\wedge\frac{1}{h_k} + \sum_{i:Y_i=1}\sum_{j:t_j\le C_i}(N_{ij}Q_i)^\wedge\left(\frac{\ddot{F}}{\dot{F}}\right)\{e^{\beta' Z_i}H(t_j)\}e^{\beta' Z_i}I(t_k\le t_j)\\ &-\sum_{i:Y_i=0}Q_i^\wedge\left(\frac{\dot{F}}{1-F}\right)\{e^{\beta' Z_i}H(C_i)\}e^{\beta' Z_i}I(t_k\le C_i),\end{aligned}$$

which gives an updating algorithm for h_k,

$$\begin{aligned}h_k = &\left\{\sum_{i:Y_i=1}(N_{ik}Q_i)^\wedge\right\}\Bigg[-\sum_{i:Y_i=1}\sum_{j:t_j\le C_i}(N_{ij}Q_i)^\wedge\left(\frac{\ddot{F}}{\dot{F}}\right)\{e^{\beta' Z_i}H(t_j)\}e^{\beta' Z_i}I(t_k\le t_j)\\ &+\sum_{i:Y_i=0}Q_i^\wedge\left(\frac{\dot{F}}{1-F}\right)\{e^{\beta' Z_i}H(C_i)\}e^{\beta' Z_i}I(t_k\le C_i)\Bigg]^{-1},\end{aligned}$$

where $\ddot{F}(x) = d^2F(x)/dx^2$. To update β, we use a one-step Newton–Raphson algorithm based on (3.3). The initial value of β is set to zero and the initial value of h_k is set to $1/K$. The E-steps and M-steps are iterated until the changes for the parameter estimates between two successive iterations are all less than 10^{-4}.

With regularity conditions, given in the Appendix, our theorems establish the asymptotic properties for the RRT estimator $\widehat{\theta}$. Denote by $\theta_0 = (\beta_0, H_0)$ the true parameter. Define the metric $d^*\{(\beta,H),(\widetilde{\beta},\widetilde{H})\} = (\|\beta-\widetilde{\beta}\|^2+\|H-\widetilde{H}\|_2^2)^{1/2}$, where $\|\cdot\|$ is the Euclidean norm and $\|H\|_2^2 = \int H(u)^2 du$.

Theorem 1 (Consistency and rate of convergence). *The RRT estimator* $(\widehat{\beta}, \widehat{H})$ *is consistent; that is,* $\widehat{\beta} \xrightarrow{P} \beta_0$ *and* $\widehat{H}(t) \xrightarrow{P} H_0(t)$ *for every* t *in the study period. The rate of convergence of* $(\widehat{\beta}, \widehat{H})$ *is of order* $n^{-1/3}$*; that is,* $d^*\{(\widehat{\beta}, \widehat{H}), (\beta_0, H_0)\} = O_p(n^{-1/3})$.

Let $w(\theta) = q\dot{F}(e^{\beta' Z}H(C))e^{\beta' Z}H(C)$, $\mathcal{E}(\theta) = qF(e^{\beta' Z}H(C)) + (1-q)c$, and $\mathcal{V}(\theta) = \mathcal{E}(\theta)\{1-\mathcal{E}(\theta)\}$. Define $m^*(\theta|O) = w(\theta)\mathcal{V}(\theta)^{-1}\{Z - g^*(C)H(C)^{-1}\}\{Y - \mathcal{E}(\theta)\}$, where

$$g^*(C) = H_0(C)\frac{E\{Zw(\theta_0)^2\mathcal{V}(\theta_0)^{-1}|C\}}{E\{w(\theta_0)^2\mathcal{V}(\theta_0)^{-1}|C\}}.$$

Theorem 2 (Asymptotic normality). *The RRT estimator* $\widehat{\beta}$ *is asymptotically normal; that is,*

$$\sqrt{n}(\widehat{\beta} - \beta_0) = \mathcal{I}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} m^*(\theta_0|O_i) + o_P(1) \xrightarrow{d} N\{0, \mathcal{I}^{-1}\Sigma(\mathcal{I}^{-1})'\},$$

where $\Sigma = E\{m^*(\theta_0|O)m^*(\theta_0|O)'\}$ *and* $\mathcal{I} = -E\{\partial m^*(\theta_0|O)/\partial\beta\}$.

Although the overall convergence rate for $(\widehat{\beta}, \widehat{H})$ is $O_P(n^{-1/3})$, it is dominated by that of the estimated nonparametric baseline function $\widehat{H}$; the convergence rate for $\widehat{\beta}$ still achieves the usual parametric rate $O_P(n^{-1/2})$. The asymptotic variance $\text{var}(\widehat{\beta}) \equiv \mathcal{I}^{-1}\Sigma(\mathcal{I}^{-1})'/n$ can be estimated by $\widehat{\text{var}}(\widehat{\beta}) \equiv \widehat{\mathcal{I}}^{-1}\widehat{\Sigma}(\widehat{\mathcal{I}}^{-1})'/n$, with $\widehat{\Sigma} = n^{-1}\sum_{i=1}^{n}\{m^*(\widehat{\theta}|O_i)m^*(\widehat{\theta}|O_i)'\}$ and $\widehat{\mathcal{I}} = -n^{-1}\sum_{i=1}^{n}\partial m^*(\widehat{\theta}|O_i)/\partial\beta$. In particular, the conditional expectations in g^* can be estimated by using the multivariate product kernel method (Huang, 1996).

4. Estimation under the Related-Question RRT

In this section, we discuss estimation with current-status data obtained from the related-question RRT (Warner, 1965). Recall that this RRT consists of a question A and a complementary question A^{c}, and that one of them, determined by a random device, is answered by the interviewee. Let Q be a binary random variable, such that $Q = 1$ represents the interviewee answering A, and $Q = 0$ represents the interviewee answering A^{c}. A single observation in the related-question RRT survey consists of $\widetilde{O} = \{X, C, Z\}$, where $X = Q\delta + (1-Q)(1-\delta)$, with $\delta = I(T \le C)$. Again, we assume that interviewees answer honestly when asked the sensitive question. Figure 1 illustrates the probability mechanism of the answer X from the related-question RRT. Below, we propose an estimation procedure for the event time model (2.1) based on the i.i.d. sample, $\widetilde{O}_i = \{X_i, C_i, Z_i\}$, for $i = 1, \ldots, n$, of the related-question RRT observation $\widetilde{O} = \{X, C, Z\}$. Here, we assume the value of $q = \Pr(Q = 1)$ is known. The case with $q = 0.5$ leads to a degenerate data distribution containing no information on model (2.1), and hence is excluded.

Let $\bar{q} = 1 - q$ and $\bar{F} = 1 - F$. Similarly to the derivation of (3.1), the likelihood function of the sample $(\widetilde{O}_1, \ldots, \widetilde{O}_n)$ takes the form

$$\widetilde{L}_n(\theta) = \prod_{i=1}^{n} \widetilde{L}(\theta|\widetilde{O}_i), \tag{4.1}$$

where $\widetilde{L}(\theta|\widetilde{O}) = [(qF + \bar{q}\bar{F})\{e^{\beta' Z} H(C)\}]^X [(1 - qF - \bar{q}\bar{F})\{e^{\beta' Z} H(C)\}]^{1-X}$. The nonparametric maximum likelihood estimator $\widehat{\theta}_r = (\widehat{\beta}_r, \widehat{H}_r)$, referred to as the rRRT estimator, of $\theta = (\beta, H)$ maximizes the likelihood (4.1) under the related-question RRT. As in Section 3, let $t_1 < \cdots < t_K$ be the ordered distinct time points of $\{C_i | X_i = 1, i = 1, \ldots, n\}$. Furthermore, as in the case of the RRT estimator $\widehat{H}$, the rRRT estimator $\widehat{H}_r$ can have jumps only at $t_j, j = 1$, for $\ldots, K$. For the computation of the rRRT estimator, we employ a similar EM algorithm to that developed in Section 3, as described below.

Let the failure time indicator $N_{ij} \equiv I(T_i \in (t_{j-1}, t_j])$, and let Q_i be the latent indicator of selecting the sensitive question in the related-question set, both of which are treated as missing values in the EM method. Let $\bar{X}_i = 1 - X_i$ and $\bar{Q}_i = 1 - Q_i$. The complete-data likelihood of the sample $\{(X_i, C_i, Z_i, N_{ij}, Q_i), i = 1, \ldots, n, j = 1, \ldots, K\}$ takes the form

$$\widetilde{L}^c_n(\theta) = \prod_{i=1}^{n} \Bigg[\Big\{ q \prod_{t_j \le C_i} \nabla F(e^{\beta' Z_i} H(t_j))^{N_{ij}} \Big\}^{Q_i X_i} \Big\{ q\bar{F}(e^{\beta' Z_i} H(C_i)) \Big\}^{Q_i \bar{X}_i}$$
$$\times \Big\{ \bar{q}\bar{F}(e^{\beta' Z_i} H(C_i)) \Big\}^{\bar{Q}_i X_i} \Big\{ \bar{q} \prod_{t_j \le C_i} \nabla F(e^{\beta' Z_i} H(t_j))^{N_{ij}} \Big\}^{\bar{Q}_i \bar{X}_i} \Bigg],$$

where $\nabla F(e^{\beta' Z_i} H(t_j)) = F(e^{\beta' Z_i} H(t_j)) - F(e^{\beta' Z_i} H(t_{j-1}))$.

By omitting terms in $\widetilde{L}^c_n(\theta)$ irrelevant to (β, H), in the M-step, we maximize

$$\sum_{i=1}^{n} \Bigg[\sum_{t_j \le C_i} X_i \{ (N_{ij} Q)^\wedge_i + \bar{X}_i (N_{ij} \bar{Q}_i)^\wedge \} \log \nabla F(e^{\beta' Z_i} H(t_j)) \tag{4.2}$$
$$+ (\bar{X}_i Q^\wedge_i + X_i \bar{Q}^\wedge_i) \log \bar{F}(e^{\beta' Z_i} H(C_i)) \Bigg],$$

where

$$Q^\wedge_i = E(Q_i | X_i, C_i, Z_i) = \Pr(Q_i = 1 | X_i, C_i, Z_i)$$
$$= \frac{q F^{X_i} \bar{F}^{1-X_i} (e^{\beta' Z_i} H(C_i))}{(qF + \bar{q}\bar{F})^{X_i} (q\bar{F} + \bar{q}F)^{1-X_i} (e^{\beta' Z_i} H(C_i))},$$
$$(N_{ij} Q_i)^\wedge = E(N_{ij} Q_i | X_i, C_i, Z_i) = \Pr(N_{ij} = 1, Q_i = 1 | X_i, C_i, Z_i)$$

$$
\begin{aligned}
&= \frac{q\nabla F(e^{\beta' Z_i} H(t_j))}{(qF + \bar{q}\bar{F})(e^{\beta' Z_i} H(C_i))} I(t_j \le C_i) X_i, \\
(N_{ij}\bar{Q}_i)^\wedge &= E\{N_{ij}(1-Q_i)|X_i, C_i, Z_i\} = \Pr(N_{ij}=1, Q_i=0|X_i, C_i, Z_i) \\
&= \frac{\bar{q}\nabla F(e^{\beta' Z_i} H(t_j))}{(q\bar{F} + \bar{q}F)\{e^{\beta' Z_i} H(C_i)\}} I(t_j \le C_i)(1-X_i),
\end{aligned}
$$

To update h_j, for $j = 1, \ldots, K$, we replace $\nabla F(e^{\beta' Z_i} H(t_j))$ in (4.2) with its first-order approximation $\dot{F}(e^{\beta' Z_i} H(t_j)) e^{\beta' Z_i} h_j$, and set the derivative of (4.2) with respect to h_k to zero to obtain an updating algorithm for h_k,

$$
\begin{aligned}
h_k = &\left[\sum_i \{X_i(N_{ik}Q_i)^\wedge + \bar{X}_i(N_{ik}\bar{Q}_i)^\wedge\}\right] \\
&\times \left[-\sum_i \sum_{j:t_j \le C_i} \{X_i(N_{ik}Q_i)^\wedge + \bar{X}_i(N_{ik}\bar{Q}_i)^\wedge\}\left(\frac{\ddot{F}}{\dot{F}}\right)\{e^{\beta' Z_i} H(t_j)\} e^{\beta' Z_i} I(t_k \le t_j)\right. \\
&\left. + \sum_i (\bar{X}_i Q_i^\wedge + X_i \bar{Q}_i^\wedge)\left(\frac{\dot{F}}{1-F}\right)\{e^{\beta' Z_i} H(C_i)\} e^{\beta' Z_i} 1(t_k \le C_i)\right]^{-1}.
\end{aligned}
$$

To update β, we use a one-step Newton–Raphson algorithm based on (4.2). We obtain the rRRT estimator by starting with initial values $\beta = 0$ and $h_j = 1/K$, for $j = 1, \ldots, K$, and then iterating between the E- and M-Steps until the changes for the parameter estimates between two successive iterations are all less than 10^{-4}.

Given the regularity conditions in the Appendix and $q \ne 0.5$, the following theorems establish the asymptotic properties for the rRRT estimator $\widehat{\theta}_r = (\widehat{\beta}_r, \widehat{H}_r)$.

Theorem 3 (Consistency and rate of convergence). *The rRRT estimator* $(\widehat{\beta}_r, \widehat{H}_r)$ *is consistent; that is,* $\widehat{\beta}_r \xrightarrow{P} \beta_0$ *and* $\widehat{H}_r(t) \xrightarrow{P} H_0(t)$ *for every* t *in the study period. The rate of convergence of* $(\widehat{\beta}_r, \widehat{H}_r)$ *is of order* $n^{-1/3}$*; that is,* $d^*\{(\widehat{\beta}_r, \widehat{H}_r), (\beta_0, H_0)\} = O_p(n^{-1/3})$.

Let $\widetilde{w}(\theta) = (q - \bar{q})\dot{F}(e^{\beta' Z} H(C)) e^{\beta' Z} H(C)$, $\widetilde{\mathcal{E}}(\theta) = (qF + \bar{q}\bar{F})\{e^{\beta' Z} H(C)\}$, and $\widetilde{\mathcal{V}}(\theta) = \widetilde{\mathcal{E}}(\theta)\{1 - \widetilde{\mathcal{E}}(\theta)\}$. Define $\widetilde{m}^*(\theta|\widetilde{O}) = \widetilde{w}(\theta)\widetilde{\mathcal{V}}(\theta)^{-1}\{Z - \widetilde{g}^*(C)H(C)^{-1}\}\{X - \widetilde{\mathcal{E}}(\theta)\}$, where

$$
\widetilde{g}^*(C) = H_0(C) \frac{E\left\{Z\widetilde{w}(\theta_0)^2 \widetilde{\mathcal{V}}(\theta_0)^{-1}|C\right\}}{E\left\{\widetilde{w}(\theta_0)^2 \widetilde{\mathcal{V}}(\theta_0)^{-1}|C\right\}}.
$$

Theorem 4 (Asymptotic normality). *The rRRT estimator* $\widehat{\beta}_r$ *is asymptotically normal; that is,*

$$\sqrt{n}(\widehat{\beta}_r - \beta_0) = \mathcal{I}_r^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \widetilde{m}^*(\theta_0|\widetilde{O}_i) + o_P(1) \xrightarrow{d} N\{0, \mathcal{I}_r^{-1}\Sigma_r(\mathcal{I}_r^{-1})'\},$$

where $\Sigma_r = E\{\widetilde{m}^*(\theta_0|\widetilde{O})\widetilde{m}^*(\theta_0|\widetilde{O})'\}$ *and* $\mathcal{I}_r = -E\{\partial\widetilde{m}^*(\theta_0|\widetilde{O})/\partial\beta\}$.

The asymptotic variance for $\widehat{\beta}_r$, $\mathrm{var}(\widehat{\beta}_r) \equiv \mathcal{I}_r^{-1}\Sigma_r(\mathcal{I}_r^{-1})'/n$, can be estimated by $\widehat{\mathrm{var}}(\widehat{\beta}_r) \equiv \widehat{\mathcal{I}}_r^{-1}\widehat{\Sigma}_r(\widehat{\mathcal{I}}_r^{-1})'/n$, where $\widehat{\Sigma}_r = n^{-1}\sum_{i=1}^{n}\{\widetilde{m}^*(\widehat{\theta}_r|\widetilde{O}_i)\widetilde{m}^*(\widehat{\theta}_r|\widetilde{O}_i)'\}$ and $\widehat{\mathcal{I}}_r = -n^{-1}\sum_{i=1}^{n}\partial\widetilde{m}^*(\widehat{\theta}_r|\widetilde{O}_i)/\partial\beta$. Based on the asymptotic normality and the estimated variance, one can perform hypothesis testing and construct a confidence interval for the regression parameter β.

The proofs for the asymptotic theory of the RRT and rRRT estimators in Theorems 1–4 are, in general, based on the techniques of Huang (1996), van der Vaart and Wellner (1996), and Korosok (2008). In short, we first apply the techniques developed in van der Vaart and Wellner (1996) to establish the consistency and the convergence rate of estimators in Theorem 1 and 3, then use the techniques similar to Huang (1996) to derive the efficient scores (m^* and $\widetilde{m}^*$) and the invertibility of the efficient Fisher Informations ($\mathcal{I}$ and $\mathcal{I}_r$), and finally follow the empirical process theory and semiparametric M-estimator theory (Korosok, 2008; van der Vaart and Wellner, 1996) to establish the asymptotic normality of the estimators in Theorems 2 and 4. Proofs for the theorems are given in the Appendix.

Remark 1. In Ahangar et al. (2012), the response probability distribution F to the sensitive issue is estimated indirectly using the probability distribution of the actual response Y (to the mix of sensitive and innocuous question problems) $\Pr(Y = 1) = qF + (1-q)c$, or by using the probability distribution of the actual response X (to the mix of sensitive and complementary question problems) $\Pr(X = 1) = qF + (1-q)(1-F)$, and the resulting estimate of F can take a value outside [0,1]. Our method differs from that of Ahangar et al. in that we directly model and estimate the distribution F, avoiding the problem of an out-of-range estimation. In the current work, we employ the same questionnaire survey designs as those of Warner (1965) and Greenberg et al. (1969), but propose a different modeling and estimation approach to those of Ahangar et al. (2012), Warner (1965), and Greenberg et al. (1969). In particular, we focus on a (type-I) interval-censored time-to-event model and data, which has not previously been addressed in the RRT literature.

5. Simulation Studies

We employ simulation studies to assess the performance of the proposed estimation methods and examine the adequacy of the associated normal approximations. Each simulation consists of 1,000 replications. The sample size $n = 500$ or 1000 for each simulated data set.

In the first simulation study, we consider the covariate vector $Z = (Z_1, Z_2)'$, where Z_1 is generated from Bernoulli(0.5), Z_2 is generated from $N(0, 0.5)$, and they are independent of each other. Given the covariate Z, the time T to the sensitive event is generated from (2.1) with $\beta = (1, 0.5)'$, $H(t) = 0.5t^{0.7}$, and $F(x) = 1 - \exp(-x)$, which corresponds to the proportional hazards (PH) model. The survey time C is simulated from Uniform(0,1), independently of T and Z. The answer to the sensitive question is given by $\delta = I(T \leq C)$, and the answer to the innocuous question W is generated as Bernoulli with $\Pr(W = 1|C, Z) = c$, with the constant $c = 0, 0.25, 0.5$, or 1. Given δ and W, the observed response Y of the unrelated-question RRT survey is $Y = Q\delta + (1 - Q)W$, and the observed response X of the related-question RRT survey is $X = Q\delta + (1 - Q)(1 - \delta)$, where Q is Bernoulli with $\Pr(Q = 1) = q$, for $q = 0.5$, 0.7, 0.9, or 1, with the case $q = 0.5$ excluded for the related-question RRT as mentioned in Section 4. The unrelated-question RRT observation consists of $O = (Y, C, Z)$, and the related-question RRT observation consists of $\widetilde{O} = (X, C, Z)$. In the second simulation study, the data are simulated from the same setups as above, except that $F(x) = x/(1 + x)$, which corresponds to the proportional odds (PO) model.

The simulation results are presented in Tables 1 and 2, including the bias and standard deviation (SD) over the simulation replicates for the estimators of β, the average standard error (ASE) based on the asymptotic theory, and the coverage probability (CP) of the 95% Wald-type confidence intervals for β over the simulation replicates. Also presented are the results of the relative efficiency (RE) of the RRT or the rRRT estimation relative to the ideal case, defined as $\text{MSE}(\widehat{\beta}_I)/\text{MSE}(\check{\beta})$, where MSE is the mean squared error over the simulation replicates, $\check{\beta}$ is the RRT estimate ($\widehat{\beta}$) or the rRRT estimate ($\widehat{\beta}_r$) of β, and $\widehat{\beta}_I$ is the "ideal" estimate obtained when the full data $\{(\delta_i, C_i, Z_i)|i = 1, \ldots, n\}$ are available, that is, the estimate from (3.1) under $q = 1$.

The results in Tables 1 and 2 show that the proposed estimators for the event time model parameters under the unrelated- and the related-question RRT surveys are essentially unbiased. Furthermore, the asymptotic theory performs well in finite samples: the asymptotic standard error approximates the simulation standard deviation well, and the Wald-type confidence interval based on asymptotic normality has a virtually correct coverage probability. As expected, the RRT estimators for both the related- and unrelated-question RRTs are less efficient than the estimator based on the current-status data without the RRT, that is, the "ideal" estimator. In addition, the relative efficiency of the RRT estimators with respect to the ideal estimator ranges from 10 to 87%, depending on the probability mechanism of the random device underlying the RRT. Comparing the results of the unrelated- and related-question RRTs, we find that the unrelated-question RRT leads to more efficient estimation than does the related-question RRT, especially when the probability c of answering "yes" to the innocuous question is smaller. The efficiency of the related-RRT estimator increases when the probability q of choosing the sensitive question instead of the

Table 1. Simulation results for the PH model under various survey designs. $(\widehat{\beta}_1, \widehat{\beta}_2)$: the unrelated-question RRT estimator; $(\widehat{\beta}_{r1}, \widehat{\beta}_{r2})$: the related-question RRT estimator; $(\widehat{\beta}_{I1}, \widehat{\beta}_{I2})$: the ideal estimator. ASE: the average of the standard error estimates; CP: the coverage probability of the 95% confidence interval; RE: the ratio of the mean squared error of the ideal estimate to that of the considered estimate.

			$n=500$					$n=1000$				
q	c		Bias	SD	ASE	CP	RE	Bias	SD	ASE	CP	RE
0.5	0	$\widehat{\beta}_1$	0.014	0.276	0.290	96.6	0.347	-0.002	0.191	0.199	96.9	0.349
		$\widehat{\beta}_2$	0.003	0.187	0.210	97.5	0.358	0.001	0.138	0.144	95.7	0.311
	0.25	$\widehat{\beta}_1$	0.075	0.358	0.362	96.7	0.198	0.029	0.238	0.244	96.6	0.222
		$\widehat{\beta}_2$	0.036	0.259	0.257	96.5	0.183	0.016	0.175	0.174	96.1	0.193
	0.5	$\widehat{\beta}_1$	0.088	0.405	0.389	97.1	0.154	0.048	0.252	0.260	96.4	0.193
		$\widehat{\beta}_2$	0.052	0.284	0.271	97.4	0.150	0.023	0.181	0.182	95.5	0.177
	1	$\widehat{\beta}_1$	0.098	0.356	0.322	93.8	0.195	0.046	0.236	0.220	94.7	0.220
		$\widehat{\beta}_2$	0.067	0.240	0.203	90.6	0.203	0.024	0.149	0.141	93.0	0.264
0.7	0	$\widehat{\beta}_1$	0.010	0.222	0.225	95.1	0.535	-0.001	0.153	0.156	95.3	0.542
		$\widehat{\beta}_2$	0.007	0.145	0.162	97.2	0.591	0.001	0.111	0.112	94.9	0.483
	0.25	$\widehat{\beta}_1$	0.037	0.249	0.246	94.4	0.417	0.015	0.165	0.170	95.6	0.461
		$\widehat{\beta}_2$	0.020	0.169	0.176	96.7	0.434	0.007	0.121	0.121	95.8	0.402
	0.5	$\widehat{\beta}_1$	0.043	0.257	0.255	95.1	0.390	0.021	0.175	0.176	95.6	0.412
		$\widehat{\beta}_2$	0.027	0.179	0.179	95.9	0.383	0.013	0.121	0.124	95.4	0.398
	1	$\widehat{\beta}_1$	0.056	0.251	0.242	95.0	0.401	0.027	0.173	0.168	94.5	0.413
		$\widehat{\beta}_2$	0.038	0.170	0.158	93.8	0.412	0.018	0.113	0.111	93.3	0.450
	-	$\widehat{\beta}_{r1}$	0.079	0.506	0.510	97.6	0.101	0.036	0.342	0.335	95.4	0.107
		$\widehat{\beta}_{r2}$	0.038	0.331	0.349	97.4	0.113	0.024	0.237	0.234	95.9	0.104
0.9	0	$\widehat{\beta}_1$	0.012	0.181	0.181	94.8	0.804	0.005	0.121	0.126	96.1	0.867
		$\widehat{\beta}_2$	0.012	0.123	0.129	96.7	0.820	0.003	0.088	0.089	95.7	0.766
	0.25	$\widehat{\beta}_1$	0.020	0.185	0.184	95.3	0.770	0.007	0.125	0.128	96.2	0.816
		$\widehat{\beta}_2$	0.016	0.127	0.130	95.9	0.759	0.005	0.091	0.090	95.1	0.722
	0.5	$\widehat{\beta}_1$	0.022	0.184	0.186	95.2	0.774	0.010	0.126	0.129	95.3	0.800
		$\widehat{\beta}_2$	0.020	0.130	0.130	96.0	0.723	0.007	0.090	0.091	95.9	0.729
	1	$\widehat{\beta}_1$	0.027	0.187	0.186	95.1	0.745	0.015	0.133	0.130	94.9	0.713
		$\widehat{\beta}_2$	0.025	0.129	0.126	94.7	0.729	0.011	0.089	0.088	95.8	0.713
	-	$\widehat{\beta}_{r1}$	0.023	0.216	0.216	94.2	0.561	0.015	0.146	0.150	96.2	0.588
		$\widehat{\beta}_{r2}$	0.023	0.148	0.152	96.5	0.561	0.009	0.105	0.105	96.2	0.534
1	-	$\widehat{\beta}_{I1}$	0.015	0.162	0.161	94.5	-	0.005	0.113	0.112	95.1	-
		$\widehat{\beta}_{I2}$	0.014	0.111	0.111	95.4	-	0.005	0.077	0.078	95.2	-

complementary question approaches one. These effects correspond to a bias–efficiency trade-off, as mentioned in Scheers and Dayton (1988); see Section 7 for further discussion.

Figure 2 shows the Q-Q plots for the standardized RRT estimates $\{\widehat{\text{var}}(\widehat{\beta})\}^{-1/2}(\widehat{\beta}-\beta)$ and the standardized rRRT estimates $\{\widehat{\text{var}}(\widehat{\beta}_r)\}^{-1/2}(\widehat{\beta}_r-\beta)$ for β with respect to the standard normal quantile values under the PH model with

Table 2. Simulation results for the PO models under various survey designs. $(\widehat{\beta}_1, \widehat{\beta}_2)$: the unrelated-question RRT estimator; $(\widehat{\beta}_{r1}, \widehat{\beta}_{r2})$: the related-question RRT estimator; $(\widehat{\beta}_{I1}, \widehat{\beta}_{I2})$: the ideal estimator. ASE: the average of the standard error estimates; CP: the coverage probability of the 95% confidence interval; RE: the ratio of the mean squared error of the ideal estimate to that of the considered estimate.

			$n = 500$					$n = 1000$				
q	c		Bias	SD	ASE	CP	RE	Bias	SD	ASE	CP	RE
0.5	0	$\widehat{\beta}_1$	-0.015	0.342	0.363	96.4	0.396	-0.018	0.233	0.249	97.0	0.406
		$\widehat{\beta}_2$	-0.011	0.229	0.262	98.0	0.413	-0.007	0.163	0.179	96.9	0.424
	0.25	$\widehat{\beta}_1$	0.062	0.456	0.466	97.0	0.220	0.024	0.305	0.316	96.7	0.237
		$\widehat{\beta}_2$	0.028	0.331	0.333	97.4	0.197	0.011	0.222	0.225	95.7	0.229
	0.5	$\widehat{\beta}_1$	0.072	0.520	0.513	96.9	0.168	0.047	0.333	0.344	96.5	0.196
		$\widehat{\beta}_2$	0.043	0.369	0.360	97.4	0.158	0.017	0.236	0.243	96.5	0.201
	1	$\widehat{\beta}_1$	0.093	0.516	0.471	95.8	0.169	0.041	0.340	0.318	94.8	0.189
		$\widehat{\beta}_2$	0.069	0.358	0.317	92.8	0.164	0.014	0.224	0.216	94.2	0.225
0.7	0	$\widehat{\beta}_1$	-0.014	0.279	0.287	94.9	0.597	-0.012	0.191	0.198	95.9	0.606
		$\widehat{\beta}_2$	-0.004	0.183	0.207	97.5	0.646	-0.006	0.138	0.143	95.6	0.590
	0.25	$\widehat{\beta}_1$	0.025	0.321	0.321	95.1	0.448	0.010	0.211	0.221	95.9	0.495
		$\widehat{\beta}_2$	0.014	0.221	0.230	96.9	0.444	0.004	0.156	0.158	95.3	0.459
	0.5	$\widehat{\beta}_1$	0.030	0.339	0.339	95.5	0.401	0.018	0.232	0.234	95.5	0.409
		$\widehat{\beta}_2$	0.022	0.240	0.241	96.7	0.376	0.010	0.161	0.166	96.0	0.437
	1	$\widehat{\beta}_1$	0.044	0.350	0.343	95.8	0.374	0.023	0.244	0.237	94.9	0.369
		$\widehat{\beta}_2$	0.032	0.241	0.238	95.5	0.367	0.014	0.168	0.164	95.2	0.399
	-	$\widehat{\beta}_{r1}$	0.077	0.672	0.692	97.8	0.102	0.031	0.435	0.446	96.8	0.116
		$\widehat{\beta}_{r2}$	0.042	0.472	0.474	98.3	0.097	0.016	0.311	0.313	96.5	0.116
0.9	0	$\widehat{\beta}_1$	-0.008	0.234	0.237	94.7	0.848	-0.006	0.158	0.164	96.3	0.881
		$\widehat{\beta}_2$	0.002	0.159	0.170	97.1	0.864	-0.003	0.116	0.118	95.3	0.836
	0.25	$\widehat{\beta}_1$	0.005	0.242	0.244	94.6	0.792	-0.001	0.165	0.169	95.8	0.812
		$\widehat{\beta}_2$	0.009	0.167	0.175	97.0	0.782	0.001	0.121	0.121	95.3	0.771
	0.5	$\widehat{\beta}_1$	0.008	0.246	0.249	94.7	0.770	0.003	0.168	0.173	95.7	0.779
		$\widehat{\beta}_2$	0.014	0.173	0.178	96.1	0.722	0.004	0.121	0.123	96.0	0.767
	1	$\widehat{\beta}_1$	0.011	0.254	0.256	95.5	0.719	0.009	0.181	0.178	95.9	0.674
		$\widehat{\beta}_2$	0.017	0.178	0.181	96.5	0.678	0.008	0.126	0.126	95.9	0.706
	-	$\widehat{\beta}_{r1}$	0.009	0.286	0.289	95.2	0.566	0.009	0.197	0.200	96.1	0.570
		$\widehat{\beta}_{r2}$	0.017	0.197	0.206	97.2	0.559	0.006	0.141	0.142	96.1	0.569
1	0	$\widehat{\beta}_{I1}$	-0.006	0.215	0.217	94.7	-	-0.007	0.149	0.151	95.8	-
		$\widehat{\beta}_{I2}$	0.002	0.148	0.155	96.9	-	-0.001	0.106	0.107	94.6	-

$q = 0.7, c = 0.25$, and $n = 1000$. The Q-Q plots confirm that the normal approximation theory for the proposed estimators is adequate.

6. Analysis of Extramarital Relations Data

In this section, we apply the proposed methods to a data set from the Taiwan Social Change Survey (TSCS) conducted by Academia Sinica in Taiwan.

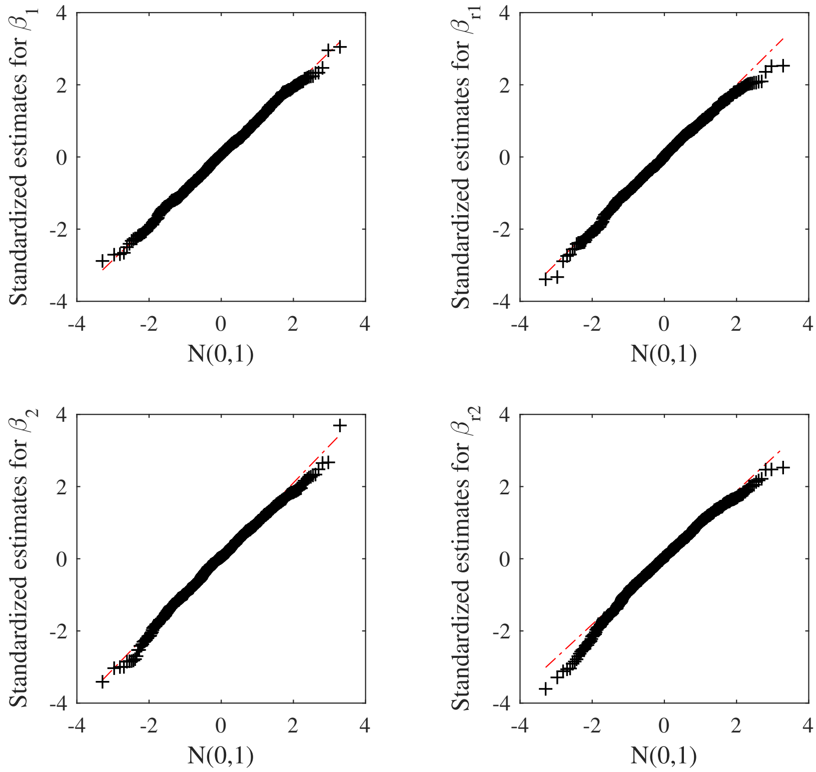

Figure 2. Q-Q plots of the standardized RRT estimates (left panels) and rRRT estimates (right panels) versus the standard normal distribution under the simulation scenario with $n = 1000, q = 0.7$, and $c = 0.25$, and the PH model.

One of the survey questions was aimed at extramarital relations, in particular, extramarital sex, and was asked using the unrelated-question RRT of Greenberg et al. (1969) for 1,140 study participants, who were aged above 18 years and married. The sensitive and innocuous questions were given by

A: *Have you ever had sex with someone other than your spouse?*
B: *Were you born in the month of January, February, or March?*

The interviewees were asked to pick up one card from the deck of 40 cards numbered 1 to 5 (8 cards numbered 1, 4 numbered 2, 8 numbered 3, 16 numbered 4, 4 numbered 5), and not to tell the interviewer the number on the card. Then, the interviewees answered Question A or B, according to the number on the card. If the number on the card was 1, 2, or 3, they answered Question A; if the number on the card was 4 or 5, they answered Question B. Under the design, the probability of answering the sensitive question A is $\Pr(Q = 1) = q = 0.5$, where Q is a binary random variable representing whether question A ($Q = 1$) or B is answered.

Table 3. Analysis results of extramarital relations data.

Model		Male	Atti	Child	EduYear	max-loglik
PH	Est	1.223*	0.791*	-0.948	-0.032	-588.128
	SE	0.320	0.312	0.505	0.038	
PO	Est	1.449*	0.950*	-1.268	-0.036	-589.355
	SE	0.373	0.433	0.673	0.048	

$n = 1040, q = 0.5, c = 0.25$, and "$*$" denotes significance.

Based on the survey data obtained using the above RRT design, we wish to study the relationship between a set of covariates and the time T to extramarital relations since marriage. Let C denote the survey time of the interviewee since he/she was married, obtained by the answer to the survey question, "How many years have you been married ?" Let $\delta = I(T \leq C)$ be the current-status indicator for the interviewee at the survey time C, representing whether or not the sensitive event (extramarital relations) had occurred by the time C. Owing to the RRT design, instead of observing T, we can only observe $Y = Q\delta + (1-Q)W$, assuming that the interviewee answered question A truthfully, and W is the response to question B.

The covariates we consider in the analysis include gender (Male; 1: Male, 0: Female), attitude toward extramarital relations (Atti; 1: Yes, 0: No), any children (Child; 1: Yes, 0: No), and years of education (EduYear; 1.5–27 years). Whether an interviewee has any children was obtained by the survey question "Do you have any child?", and the attitude toward extramarital relations of an interviewee was obtained by the survey question "Can married people have extramarital relations?" To investigate the covariate effects on the time to extramarital relations, we consider the class of generalized odds-rate hazards models (Scharfstein, Tsiatis and Gilbert, 1998), that is, the function F in (2.1) is given by $F(x) = 1 - (1 + \nu x)^{1/\nu}, \nu \geq 0$. This class of models includes the PH ($\nu = 0$) and PO ($\nu = 1$) models as special cases. For the TSCS RRT data, the generalized odds-rate hazards model with $\nu = 0$, that is, the PH model is found to give the largest log-likelihood value (-558.128) over a set of grid values for $\nu \in [0, 2]$. The regression analysis results for the TSCS RRT data based on the PH and PO models are presented in Table 3.

As shown in Table 3, the results from both models reveal that males have a significantly higher cumulative chance of experiencing extramarital relations than females do. Furthermore, persons with a positive attitude toward extramarital relations have a significantly shorter time to extramarital relations than persons with a negative attitude do. On the other hand, people with children and who have more education tend to have less incidence of extramarital relations, although these trends are not statistically significant at the 5% level.

Figure 3 presents the RRT estimates of the cumulative percentages of extramarital relations over years since marriage for males and females, obtained

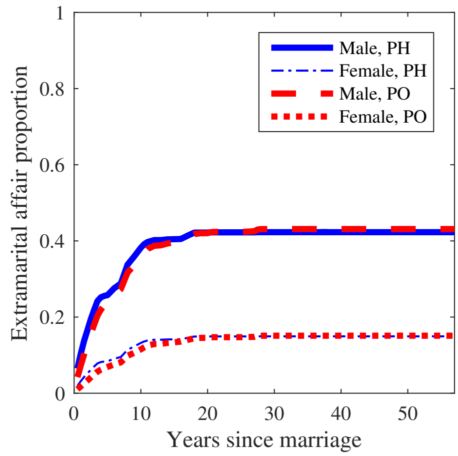

Figure 3. The estimated cumulative proportions of extramarital relations for males and females under the PH model and the PO model, with the covariates (Atti, Child, EduYear) fixed at the sample means.

from the PH and PO models, with the covariates Atti, Child, and EduYear fixed at their sample means. The estimates of the proportions of males and females having extramarital relations based on Figure 3 are presented in the Supplementary Material.

Here, to help readers understand the proposed method, we summarize and explain the steps applied to the TSCS unrelated-question RRT data. First, we set the time-to-event model, which can belong to a general class of models, such as the generalized odds-rate hazard models. The resulting likelihood is then obtained from (3.1) of Section 3. Second, we apply the computation algorithm described in Section 3 to maximize the likelihood function and obtain the maximum likelihood estimates of the model parameters. Third, we base on the asymptotic normality theory in Theorem 2 to make the inferences, including the hypothesis testing and confidence intervals, about model parameters. Fourth, we choose an appropriate time-to-event model from the general class of models considered, and compare the log-likelihood values of different models in the class. When the method is applied to related-question RRT data, the steps are essentially the same, except that the likelihood function is now obtained from expression (3.1) of Section 4, and the computation algorithm and the asymptotic normality theory of the maximum likelihood estimates are provided in Section 4 and Theorem 4, respectively.

7. Conclusion

Current-status data obtained from routine survey studies, where continuous follow-up is rare, allow us to analyze the time-to-event distribution and the distribution conditional on certain covariate variables (Huang, 1996; Jewell and van der Laan, 2003). In this work, we extend the regression analysis for current-status data to that for data collected using theRRT, including the unrelated-question RRT of Greenberg et al. (1969) and the related-question RRT of Warner (1965). The RRT is usually applied in surveys on sensitive issues, and uses some random perturbation of the target question in a well-designed manner to make it possible to analyze the target issue while protecting interviewees' privacy. Statistical analysis methods for the prevalence of a sensitive event, possibly adjusting for certain covariates, based on RRT data are available. The main and novel contribution of this work is that we extend the application of the RRT to event time analysis.

Our simulation studies reveal that the unrelated-question RRT leads to a more efficient regression coefficient estimation than that of the related-question RRT. In particular, the unrelated-question RRT achieves higher efficiency when the probability c of answering "yes" to the innocuous question is smaller. However, a smaller value of c would make the group of persons with the sensitive event easier to identify, which, in turn, would make that group of persons less willing to answer the sensitive question truthfully, leading to larger bias. In particular, the extreme case of $c = 0$ would not be used in practice, because it would make the group of persons with the sensitive event fully identifiable. On the other hand, increasing the probability q of selecting the sensitive question, rather than the innocuous question (in the unrelated-question RRT) or the complement of the sensitive question (in the related-question RRT), can yiled higher estimation efficiency; and the extreme case of $q = 1$ yields the current-status data exactly for the sensitive event, and hence full efficiency. However,a larger value of q means less privacy protection, and thus may induce larger bias, as above. Hence, there is a bias–efficiency trade-off when analyzing survey data on sensitive issues.

In this work, we assume that interviewees answer the sensitive question truthfully under the RRT. Relaxing this assumption requires an extension of the methods proposed here for event time regression analysis, and so is left as poissible.

Supplementary Materials

The Matlab code implementing the proposed methods and a more detailed report of the TSCS data analysis are available in the online Supplementary Material.

Acknowledgments

The authors thank the associate editor and two referees for their helpful comments, and the Ministry of Science and Technology of Republic of China for its support.

Appendix

We use the notation $\mathbb{P}_n, P_0$, and P for the expectations taken under the empirical distribution, the true underlying distribution, and a given model, respectively. Assume Z is d-dimensional. The parameter space of β is a compact subset $\mathcal{B}$ of $\mathbb{R}^d$, and the parameter space of H is a set $\mathcal{H}$ of all right-continuous non-decreasing functions that are uniformly bounded on the study period $[0, \tau]$. The asymptotic theories are based on the following assumptions.

(C1) The covariate vector Z is bounded with $E\{\text{var}(Z|C)\}$ is positive definite.

(C2) The density of survey time C is continuous with support $[\tau_1, \tau_2]$ where $0 < \tau_1 < \tau_2 < \tau$.

(C3) The true parameter β_0 is an interior point of its parameter space, and H_0 is continuously differentiable and satisfies $M^{-1} < H_0(\tau_1) < H_0(\tau_2) < M$.

Our proofs of the theorems are mainly based on the techniques developed in Huang (1996), van der Vaart (1998), and Korosok (2008). Specifically, we apply Theorem 5.7 of van der Vaart (1998) and Theorem 3.2.5 of van der Vaart and Wellner (1996) to establish the consistency and the convergence rate of estimators. To derive the asymptotic normality of estimators, we first derive the efficient score for β using techniques similar to Huang (1996), and then follow the well-known empirical process theory and semiparametric M-estimator theory (e.g., Korosok (2008), Korosok (2008), van der Vaart and Wellner (1996)) to obtain the asymptotic normality of estimators. Without confusion, we also write $L(\theta) = L(\theta|O), m^*(\theta) = m^*(\theta|O), \widetilde{L}(\theta) = \widetilde{L}(\theta|\widetilde{O})$, and $\widetilde{m}^*(\theta) = \widetilde{m}^*(\theta|\widetilde{O})$.

A.1. Asymptotic properties of the RRT estimator

Proof of Theorem 1.

Consistency. First, we apply Theorem 5.7 of van der Vaart (1998) to establish the consistency of RRT estimator $(\widehat{\beta}, \widehat{H})$. Let $m(\theta) = \log L(\theta)$. Since the class of monotone and uniformly bounded functions is a Donsker class, by Theorem 2.10.6 of van der Vaart and Wellner (1996) and conditions (C1)-(C3), we know that the class $\{m(\theta)|\theta \in \mathcal{B} \times \mathcal{H}\}$ is Donsker and hence Glivenko-Cantelli. By Jensen's inequality, we have

$$P_0\{m(\theta) - m(\theta_0)\} \leq \log\left[P_0\left\{\frac{L(\theta)}{L(\theta_0)}\right\}\right] = 0,$$

wherein the equality holds only if $L(\theta) = L(\theta_0)$ a.s., or equivalently, $\theta = \theta_0$ by model identifiability. This indicates that

$$\sup_{\{\theta:\ d^*\{\theta,\theta_0\}>\varepsilon\}} P_0 m(\theta) < P_0 m(\theta_0).$$

Furthermore, by definition of $\widehat{\theta}$, $\mathbb{P}_n m(\theta_0) \le \mathbb{P}_n m(\widehat{\theta})$. Applying Theorem 5.7 of van der Vaart (1998), we have $\widehat{\beta}\to\beta_0$ and $\|\widehat{H}-H_0\|_2\to 0$ in probability. Since H_0 is continuous and strictly monotone, it further implies $\widehat{H}(t)\to H_0(t)$ in probability for every t in (τ_1, τ_2).

Rate of convergence. Below we apply Theorem 3.2.5 of van der Vaart and Wellner (1996) to prove $d^*(\widehat{\theta}, \theta_0) = O_p(n^{-1/3})$. Here we follow van der Vaart (1998) to introduce the bracketing number and bracketing integral. For two functions l and u, define the bracket $[l, u]$ as the set of all functions f with $l \le f \le u$. An ε-bracket in $L_2(P) = \{f : Pf^2 < \infty\}$ with respect to some distribution P is then a bracket $[l, u]$ with $P(u-l)^2 < \varepsilon^2$. For a subclass $\mathcal{C}$ of $L^2(P)$, the bracketing number $N_{[\,]}(\varepsilon, \mathcal{C}, L_2(P))$ is defined as the minimum number of ε-bracket that is needed to cover $\mathcal{C}$, and the bracketing integral $J(\delta, \mathcal{C}, L_2(P))$ is defined as $\int_0^\delta \{1 + \log N_{[\,]}(\varepsilon, \mathcal{C}, L_2(P))\}^{1/2} d\varepsilon$. With the consistency and (C3), we restrict H to $\mathcal{H}_0 = \{H \in \mathcal{H} \mid M^{-1} \le H(\tau_1) \le H(\tau_2) \le M\}$. Let $\Psi = \{m(\theta) : \theta \in \mathcal{B} \times \mathcal{H}_0\}$. Then each element in Ψ is uniformly bounded and satisfies $P_0\{m(\theta) - m(\theta_0)\}^2 \preceq d^*\{\theta, \theta_0\}^2$ by the mean value theorem, where $\preceq$ means smaller than, up to a constant. Lemma A.1 below gives the bracketing integral $J\{\delta, \Psi, L_2(P)\} = O(\delta^{1/2})$. Consequently, Lemma 19.36 of van der Vaart and Wellner (1996) gives

$$P^* \sup_{d^*(\zeta,\zeta_0)<\delta} |\sqrt{n}(\mathbb{P}_n - P_0)\{m(\theta) - m(\theta_0)\}| \preceq \delta^{1/2}\left(1 + \frac{\delta^{1/2}}{\delta^2\sqrt{n}}\right),$$

where P^* is the outer expectation. By the inequality of Kullback-Leibler divergence (van der Vaart and Wellner, 1996, p.62), $E\{m(\theta|O)|C, Z\}$ is maximized at $(\theta_0, H_0(C))$. So, its first derivative is equal to zero there. Since (C, Z) has a bounded support, the parameter spaces are compact, and H is uniformly bounded from 0 and ∞, a Taylors's expansion gives $P_0\{m(\theta) - m(\theta_0)\} \preceq -d^*(\theta, \theta_0)^2$. According to Theorem 3.2.5 of van der Vaart and Wellner (1996), we complete the proof.

Lemma 1. $\log N_{[\,]}(\varepsilon, \Psi, L_2(P_0)) = O(1/\varepsilon)$.

Proof. First consider the functions in Ψ for a fixed β. Given the ε-brackets $H^L \le H \le H^U$, it is readily to get a bracket (m^L, m^U) for $m(\theta)$ where

$$m^L \equiv \log\left[\{qF(e^{\beta' Z}H^L(C)) + (1-q)c\}^Y\{1 - qF(e^{\beta' Z}H^U(C)) - (1-q)c\}^{1-Y}\right],$$

$$m^U \equiv \log\left[\{qF(e^{\beta' Z}H^U(C)) + (1-q)c\}^Y\{1 - qF(e^{\beta' Z}H^L(C)) - (1-q)c\}^{1-Y}\right].$$

By the mean value theorem, we have $|m^L - m^U|^2 \preceq \{H^U(C) - H^L(C)\}^2$. Thus, brackets for H of $\|\cdot\|_2$-size ε can translate into brackets for $m(\theta)$ of $L_2(P_0)$-size proportional to ε. By Example 19.11 of van der Vaart (1998), we can cover the set of all H by $\exp(C/\varepsilon)$ brackets of size ε for some constant C. Next we allow β to vary freely as well. Because β is finite-dimensional and $(\partial/\partial\beta)m(\theta|O)$ is uniformly bounded in (θ, O), this increases the entropy only slightly. This completes the proof.

Proof of Theorem 2.
Efficient score. Recall that $w(\theta) = q\dot{F}(e^{\beta' Z}H(C))e^{\beta' Z}H(C), \mathcal{E}(\theta) = qF(e^{\beta' Z} H(C)) + (1-q)c$, and $\mathcal{V}(\theta) = \mathcal{E}(\theta)\{1-\mathcal{E}(\theta)\}$. The score for β, defined by $\partial m/\partial\beta$, takes the form $m_1(\theta) = Zw(\theta)\mathcal{V}(\theta)^{-1}\{Y - \mathcal{E}(\theta)\}$. Consider parametric paths $H_\varepsilon \in \mathcal{H}$ with $H_\varepsilon|_{\varepsilon=0} = H$ and $(\partial H_\varepsilon/\partial\varepsilon)|_{\varepsilon=0} = g$. The score for H, defined by $\{\partial m(\beta, H_\varepsilon)/\partial\varepsilon\}|_{\varepsilon=0}$, has the form $m_2(\theta)[g] = \{g(C)/H(C)\}w(\theta)\mathcal{V}(\theta)^{-1} \{Y - \mathcal{E}(\theta)\}$. Also define $m_{12}(\theta)[g] = \{\partial m_1(\beta, H_\varepsilon)/\partial\varepsilon\}|_{\varepsilon=0}$ and $m_{22}(\theta)[\tilde{g}, g] = \{\partial m_2(\beta, H_\varepsilon)[\tilde{g}]/\partial\varepsilon\}|_{\varepsilon=0}$. They have forms

$$m_{12}(\theta)[g] = -Z\frac{g(C)}{H(C)}\frac{w(\theta)^2}{\mathcal{V}(\theta)}, \quad m_{22}(\theta)[\tilde{g}, g] = -\frac{\tilde{g}(C)}{H(C)}\frac{g(C)}{H(C)}\frac{w(\theta)^2}{\mathcal{V}(\theta)}.$$

Following semiparametric M-estimator (e.g., Korosok (2008)), the efficient score is defined by $m^*(\theta) = m_1(\theta) - m_2(\theta)[g^*]$, where g^* satisfies that

$$P_0\{m_{12}(\theta_0)[g] - m_{22}(\theta_0)[g^*, g]\} = 0, \tag{A.2}$$

for all g in $L_2(P_0)$. It immediately gives that

$$g^*(C) = H_0(C)\frac{E\left[Zw(\theta_0)^2\mathcal{V}(\theta_0)^{-1}|C\right]}{E\left[w(\theta_0)^2\mathcal{V}(\theta_0)^{-1}|C\right]},$$

and hence $m^*(\theta) = w(\theta)\mathcal{V}(\theta)^{-1}\left\{Z - g^*(C)H(C)^{-1}\right\}\left\{Y - \mathcal{E}(\theta)\right\}$.

Asymptotic Normality. Denote $a^{\otimes 2} = aa'$ for any column vector a. By direct calculations and the definition of g^*, we have

$$\begin{aligned} -P_0\left\{\frac{\partial}{\partial\beta}m^*(\theta_0)\right\} &= E\left\{Z\left(Z - \frac{g^*(C)}{H_0(C)}\right)'\frac{w(\theta_0)^2}{\mathcal{V}(\theta_0)}\right\} \\ &= E\left\{\left(Z - \frac{g^*(C)}{H_0(C)}\right)^{\otimes 2}\frac{w(\theta_0)^2}{\mathcal{V}(\theta_0)}\right\}, \end{aligned}$$

which is positive definite. This implies the invertibility of $\mathcal{I}$.

Applying Taylor expansions of $m^*(\beta_0, H(C))(O)$ at $H_0(C)$, we conclude that

$$\begin{aligned} &P_0 m^*(\beta_0, H) \\ &= P_0 m^*(\theta_0) + P_0\{m_{12}(\theta_0)[H - H_0] - m_{22}(\theta_0)[g^*, H - H_0]\} + O_p(\|H - H_0\|_2^2). \end{aligned}$$

Using the facts that $P_0 m^*(\theta_0) = 0$, (A.2), and the rate of convergence of $\widehat{H}$, we have

$$\sqrt{n}P_0 m^*(\beta_0, \widehat{H}) = o_p(1). \tag{A.3}$$

It is known that the class of uniformly bounded functions of bounded variations is a Donsker class. By Theorem 2.10.6 of van der Vaart and Wellner (1996), we can show that $\{m^*(\theta)|\theta \in \mathcal{B} \times \mathcal{H}_0\}$ is a uniformly bounded Donsker class; the proof of which is technical and hence omitted here. This together with the consistency of $\widehat{\theta}$ implies that $\sqrt{n}(\mathbb{P}_n - P_0)\{m^*(\widehat{\theta}) - m^*(\theta_0)\} = o_p(1)$. Adding (A.3) and using the fact that $P_0 m^*(\theta_0) = \mathbb{P}_n m^*(\widehat{\theta}) = 0$, we obtain

$$-\sqrt{n}P_0\{m^*(\widehat{\theta}) - m^*(\beta_0, \widehat{H})\} = \sqrt{n}\mathbb{P}_n m^*(\theta_0) + o_P(1).$$

By the mean value theorem, there exists $\tilde{\beta}$ lying between $\widehat{\beta}$ and β_0 such that

$$-\sqrt{n}P_0\left\{\frac{\partial}{\partial\beta} m^*(\tilde{\beta}, \widehat{H})\right\}(\widehat{\beta} - \beta_0) = \sqrt{n}\mathbb{P}_n m^*(\theta_0) + o_P(1).$$

By the consistency of $\widehat{\theta}$ and the invertibility of $\mathcal{I}$, we have

$$\sqrt{n}(\widehat{\theta} - \theta_0) = I_0^{-1}\sqrt{n}\mathbb{P}_n m^*(\theta_0) + o_P(1) \xrightarrow{d} N(0, I^{-1}\Sigma(I^{-1})').$$

This completes the proof.

A.2. Asymptotic properties of the rRRT estimator

Proof of Theorem 3. Let $\widetilde{m}(\theta) = \log \tilde{L}(\theta)$. The consistency and the rate of convergence of the rRRT estimator can be obtained using the same argument for establishing Theorem 1 for the RRT estimator. We skip the most similar part of the proof except that the construction of the bracket for $\widetilde{m}(\theta)$ needed in Lemma A.1. Recall that

$$\widetilde{m}(\theta) = X \log\left[(qF + \bar{q}\bar{F})\{e^{\beta' Z}H(C)\}\right] + \bar{X}\log\left[(\bar{q}F + q\bar{F})\{e^{\beta' Z}H(C)\}\right].$$

Given a ε-brackets $H^L \le H \le H^U$, a bracket (m^L, m^U) for $\widetilde{m}$ can be obtained by

$$\begin{aligned} m^L(\theta) = {} & X\log\left\{qF(e^{\beta' Z}H^L(C)) + \bar{q}\bar{F}(e^{\beta' Z}H^U(C))\right\} \\ & + \bar{X}\log\left\{\bar{q}F(e^{\beta' Z}H^L(C)) + q\bar{F}(e^{\beta' Z}H^U(C))\right\}, \\ m^U(\theta) = {} & X\log\left\{qF(e^{\beta' Z}H^U(C)) + \bar{q}\bar{F}(e^{\beta' Z}H^L(C))\right\} \\ & + \bar{X}\log\left\{\bar{q}F(e^{\beta' Z}H^U(C)) + q\bar{F}(e^{\beta' Z}H^L(C))\right\}. \end{aligned}$$

Proof of Theorem 4. The asymptotic normality of the rRRT estimator can be obtained using the same argument in Theorem 2. Below we derive the efficient score $\widetilde{m}^*(\theta)$ and show the invertibility of efficient information $\mathcal{I}_r$, but skip the much similar remaining part of the proof.

Recall that $\widetilde{w}(\theta)=(q-\bar{q})\dot{F}(e^{\beta' Z}H(C))e^{\beta' Z}H(C)$, $\widetilde{\mathcal{E}}(\theta)=(qF+\bar{q}\bar{F})\{e^{\beta' Z}H(C)\}$, and $\widetilde{\mathcal{V}}(\theta)=\widetilde{\mathcal{E}}(\theta)\{1-\widetilde{\mathcal{E}}(\theta)\}$. The score for β takes the form

$$\widetilde{m}_1(\theta)=Z\widetilde{w}(\theta)\widetilde{\mathcal{V}}(\theta)^{-1}\{X-\widetilde{\mathcal{E}}(\theta)\},$$

and the score for H along direction g takes the form

$$\widetilde{m}_2(\theta)[g]=\frac{g(C)}{H(C)}\widetilde{w}(\theta)\widetilde{\mathcal{V}}(\theta)^{-1}\{X-\widetilde{\mathcal{E}}(\theta)\}.$$

Define $\widetilde{m}_{12}(\theta)[g]=\{\partial\widetilde{m}_1(\beta,H_\varepsilon)/\partial\varepsilon\}|_{\varepsilon=0}$ and $\widetilde{m}_{22}(\theta)[\tilde{g},g]=\{\partial\widetilde{m}_2(\beta,H_\varepsilon)[\tilde{g}]/\partial\varepsilon\}|_{\varepsilon=0}$, some calculations give

$$\widetilde{m}_{12}(\theta)[g]=-Z\frac{g(C)}{H(C)}\frac{\widetilde{w}(\theta)^2}{\widetilde{\mathcal{V}}(\theta)},\quad \widetilde{m}_{22}(\theta)[\tilde{g},g]=-\frac{\tilde{g}(C)}{H(C)}\frac{g(C)}{H(C)}\frac{\widetilde{w}(\theta)^2}{\widetilde{\mathcal{V}}(\theta)}.$$

The efficient score for β is $\widetilde{m}^*(\theta)=\widetilde{m}_1(\theta)-\widetilde{m}_2(\theta)[\widetilde{g}^*]$ with $\widetilde{g}^*$ satisfying

$$P_0\{\widetilde{m}_{12}(\theta_0)[g]-\widetilde{m}_{22}(\theta_0)[\widetilde{g}^*,g]\}=0.$$

It immediately gives that

$$\widetilde{g}^*(C)=H_0(C)\frac{E\left[Z\widetilde{w}(\theta_0)^2\widetilde{\mathcal{V}}(\theta_0)^{-1}|C\right]}{E\left[\widetilde{w}(\theta_0)^2\widetilde{\mathcal{V}}(\theta_0)^{-1}|C\right]},$$

and hence $\widetilde{m}^*(\theta)=\widetilde{w}(\theta)\widetilde{\mathcal{V}}(\theta)^{-1}\{Z-\widetilde{g}^*(C)H(C)^{-1}\}\{X-\widetilde{\mathcal{E}}(\theta)\}$. By direct calculations and the definition of $\widetilde{g}^*$, we have

$$\begin{aligned}-P_0\left\{\frac{\partial}{\partial\beta}\widetilde{m}^*(\theta_0)\right\}&=E\left[Z\left\{Z-\frac{\widetilde{g}^*(C)}{H_0(C)}\right\}'\frac{\widetilde{w}(\theta_0)^2}{\widetilde{\mathcal{V}}(\theta_0)}\right]\\&=E\left[\left\{Z-\frac{\widetilde{g}^*(C)}{H_0(C)}\right\}^{\otimes 2}\frac{\widetilde{w}(\theta_0)^2}{\widetilde{\mathcal{V}}(\theta_0)}\right],\end{aligned}$$

which is positive definite. This implies the invertibility of $\mathcal{I}_r$.

References

Ahangar, R., Wang, R., Perez, J. and Singh, S. (2012). Extensive study of logistic regression using randomized response sampling. *AMSE Journals* **33**, 1–26.

Gjestvang C. R. and Singh S. (2006). A new randomized response model. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **68**, 523–530.

Greenberg, B. G., Abul-Ela, A. L., Simmons, W. R. and Horvitz, D. G. (1969). The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association* **64**, 520–539.

Horvitz, D. G., Shah, B. V. and Simmons, W. R. (1967). The unrelated question randomized response model. In *Proceedings of the Annual Meeting of the Social Statistics Section*, 65–72. American Statistical Association, Washington, D. C.

Huang, J. (1996). Efficient estimation for the Cox model with interval censoring. *The Annals of Statistics* **24,** 540–568.

Lin, D. Y., Oakes D. and Ying Z. (1998). Additive hazards regression with current status data. *Biometrika* **85,** 289–298.

Jewell, N. P. and van der Laan, M. (2003) Current status data: Review, recent developments and open problems. *Handbook in Statistics* **23**, 625–642.

Korosok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* Springer, New York.

Kuk A. Y. C. (1990). Asking sensitive questions indirectly. *Biometrika* **77**, 436–438.

Martinussen T. and Scheike T. H. (2002). Efficient estimation in additive hazards regression with current status data. *Biometrika* **89**, 649–658.

Narjis, G. and Shabbir, J. (2023) An improved two-stage randomized response model for estimating the proportion of sensitive attribute. *Sociological Methods & Research* **52**, 335–355.

Scharfstein, D. O., Tsiatis, A. A. and Gilbert, P. B. (1998). Semiparametric efficient estimation in the generalized odds-rate class of regression models for right-censored time-to-event data. *Lifetime Data Analysis* **4**, 355–391.

Scheers, N. J. and Dayton, C. M. (1988). Covariate randomized response models. *Journal of the American Statistical Association* **83**, 969–974.

Singh S., Singh R. and Mangat N. S. (2000). Some alternative strategies to Moor's model in randomized response sampling. *Journal of the Statistical and Planing Inference* **83**, 243-255.

Sun J. (2006). *The Statistical Analysis of Interval-Censored Failure Time Data.* New York, Springer-Verlag.

Sun J. and Sun L. (2005). Semiparametric linear transformation models for current status data. *The Canadian Journal of Statistics* **33**, 85–96.

Tian L. and Cai T. (2006). On the accelerated failure time model for current status and interval censored data. *Biometrika* **93**, 329–342.

Turnbull, B. W. (1976). The empirical distribution function with arbitrarily grouped, censored and truncated Data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **38**, 290–295.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge University Press, Cambridge.

van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer-Verlag, New York.

Warner S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* **60**, 63–69.

Wen C. C. and Chen Y. H. (2012). Conditional score approach to errors-in-variable current status data under the proportional odds model. *Scandinavian Journal of Statistics* **39**, 635–644.

Zeng, D. and Lin, D. Y. (2006). Efficient estimation of semiparametric transformation models for counting processes. *Biometrika* **93**, 627–640.

Zeng, D. and Lin, D. Y. (2007). Maximum likelihood estimation in semiparametric regression models with censored data. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **69**, 507–564.

Zhang Z. and Sun J. (2010). Interval censoring. *Statistical Methods in Medical Research* **19**, 53–70.

(Received September 2022; accepted April 2023)