

Non-randomized response model for sensitive survey with noncompliance

Qin Wu¹ and Man-Lai Tang²

Statistical Methods in Medical Research

0(0) 1–13

© The Author(s) 2014

Reprints and permissions:

sagepub.co.uk/journalsPermissions.nav

DOI: 10.1177/0962280214533022

smm.sagepub.com



Abstract

Collecting representative data on sensitive issues has long been problematic and challenging in public health prevalence investigation (e.g. non-suicidal self-injury), medical research (e.g. drug habits), social issue studies (e.g. history of child abuse), and their interdisciplinary studies (e.g. premarital sexual intercourse). Alternative data collection techniques that can be adopted to study sensitive questions validly become more important and necessary. As an alternative to the famous Warner randomized response model, non-randomized response triangular model has recently been developed to encourage participants to provide truthful responses in surveys involving sensitive questions. Unfortunately, both randomized and non-randomized response models could underestimate the proportion of subjects with the sensitive characteristic as some respondents do not believe that these techniques can protect their anonymity. As a result, some authors hypothesized that lack of trust and noncompliance should be highest among those who have the most to lose and the least to use for the anonymity provided by using these techniques. Some researchers noticed the existence of noncompliance and proposed new models to measure noncompliance in order to get reliable information. However, all proposed methods were based on randomized response models which require randomizing devices, restrict the survey to only face-to-face interview and are lack of reproducibility. Taking the noncompliance into consideration, we introduce new non-randomized response techniques in which no covariate is required. Asymptotic properties of the proposed estimates for sensitive characteristic as well as noncompliance probabilities are developed. Our proposed techniques are empirically shown to yield accurate estimates for both sensitive and noncompliance probabilities. A real example about premarital sex among university students is used to demonstrate our methodologies.

Keywords

noncompliance, non-randomized response technique, sensitive question

¹Department of Statistics, School of Mathematical Sciences, South China Normal University, Guangzhou, PR China

²Department of Mathematics and Statistics, Hang Seng Management College, Shatin, N.T., Hong Kong, PR China

Corresponding author:

Man-Lai Tang, Department of Mathematics and Statistics, Hang Seng Management College, Shatin, N.T., Hong Kong, PR China.

Email: mltang@hsmc.edu.hk

I Introduction

The growing high demand of health and medical interest in sensitive attribute (e.g. drug and alcohol use/abuse, embezzlement, high-risk sexual practices, religious preferences, sexual tendencies, racial attitudes) makes data collection and analysis of sensitive question important and necessary since efficient strategies can be well developed to satisfy the need of our society. Gaining valid answers to sensitive questions, however, has long been an age-old problem in survey research as false replies might be caused by self-protection, or refusal of participation might be caused by embarrassing nature of the questions. Most seriously, both response and non-response biases limit the ability to draw valid inference about the target population.

Various techniques have been developed to maximize cooperation, minimize the respondent's feeling of jeopardy, and guarantee anonymity. Three major techniques for these purposes include the randomized response technique (RRT),¹ unmatched count technique (UCT),² and non-randomized response technique (NRRT).³ The crux of the original Warner RRT and all its variations⁴⁻⁶ is that the real status of respondent's answer is hidden by a deliberate contamination of the data via a randomizing device. However, as pointed out by Böckenholt and Van der Heijden,⁷ RRTs are not often applied since: (i) their efficiency is low; (ii) they may not be easily followed by every respondent despite their privacy protection mechanisms; and (iii) aggregate-level estimate is not linked to individual-level covariates. To address the efficiency and aggregation issues, a class of so-called item randomized-response (IRR) models (in which a person parameter is estimated based on multiple measures of the sensitive behavior under study) was introduced. Nonetheless, drawbacks such as higher cost and lack of reproductivity induced by the introduction of the randomizing device still exist. Unlike RRTs, UCTs and NRRTs protect respondents' privacy by a deliberate contamination of the data via one or more innocuous/non-sensitive items/questions. A fatal drawback of UCT and its refinement that their advocates usually overlook is that the privacy is no longer protected if respondents in the so-called sensitive question group wish to answer all sensitive and non-sensitive questions affirmatively. If this issue occurs to many respondents, then the validity of the survey measurement may be compromised. Increasing the number of non-sensitive items may be a possible solution; however, the resultant estimator will be statistically inefficient.

The NRRT is first introduced by Tian et al.³ and Yu et al.⁸ Basically, there are two NRRTs, namely the triangular and crosswise models. In particular, the NRRT based on triangular model outperforms the other techniques as it (i) extends the applicability of RRT which is always restricted by its usage of randomizing device; (ii) is shown to be generally more efficient than its crosswise model which is in fact the non-randomized version of the well-known Warner's RR model; (iii) never reveals the true status of those participants who possess the sensitive characteristic, which UCT may not guarantee; and (iv) belongs to the class of the admissible design defined by Nayak.⁹ The NRRTs have received increased attention recently.^{10,11} Due to the aforementioned advantages, we will focus our discussion solely on the non-randomized response triangular model (NRRTM) in this article.

Most of the existing models for sensitive attributes assume that respondents will provide true answers and comply with the survey design. However, it has been reported in various experimental studies that some respondents still chose those answers which could demonstrate their positive image even being told that answers could not leak their privacy.^{12,13} This kind of noncompliance could, however, lead to severe response bias. For instance, Edgell et al.¹² reported that about 25% of the respondents did not follow the instructions when answering a question on homosexual experience and the parameter of interest was underestimated due to noncompliance. In another study (see Section 5 for detail) of premarital sex among college students in Changchun of Jilin Province, China, the NRRTM reported that around 19% of the respondents have ever had premarital sexual intercourse, which is fairly consistent with the result reported in a recent cross-sectional

study using anonymous self-questionnaire among Beijing college students.¹⁴ On the other hand, China Health and Family Life Survey (CHFLS), a national probability survey of sexual behavior conducted between 1999 and 2000, reported that almost 36% of never-married adults aged 20–34 have ever had sex.¹⁵ The striking difference between the two proportions could be partly attributed to different (i) target populations (i.e. college students vs. never-married adults aged 20–34); (ii) scales (i.e. regional vs. national); (iii) periods (i.e. 2012 vs. 2000) and (iv) survey techniques (i.e. NRRT vs. interview). Here, we hypothesize that the true proportion of college students having premarital sex is substantially underestimated by noncompliance. Indeed, after incorporating noncompliance into our NRRTM, the proportion is estimated to be 32%. Obviously, one must take the noncompliance into consideration to draw reliable conclusion.

Recently, some researchers have taken the noncompliance in surveys into account in order to draw accurate and reasonable statistical inference. For instance, by introducing mixture components in their proposed IRR models, Böckenholt and Van der Heijden⁷ developed mixture versions of the IRR models that allow for respondents who do not follow the randomized response instructions. Cruyff et al.¹⁶ developed a log-linear model which includes a so-called self protection (SP) parameter that accounts for self-protective response behavior for randomized response data (e.g. Böckenholt and Van der Heijden,⁷ Böckenholt et al.,¹⁷ and Cruyff et al.¹⁶). These models focus on the multiple sensitive questions and relate the estimates with the individual covariates. However, randomizing devices are still required in these models and they cause the non-reproductivity issue (i.e. the same respondent may provide different answers for different trials as the answers partly depend on the outcomes of the randomizing device). Besides, the usage of randomizing device also increases the cost and restricts the survey to face-to-face interview. As a result, simple models for measuring the noncompliance are very necessary and important in practice. Two new NRRTMs with their point estimates and confidence intervals are proposed under noncompliance in Section 2. Their sample size formulae are also developed in Section 3. In Section 4, a simulation study is conducted to evaluate the performance of the two models. The aforementioned premarital sex study is adopted to demonstrate our methods in Section 5. A brief conclusion is presented in Section 6.

2 NRRTMs with noncompliance

2.1 The original NRRTM

Originally, two non-randomized response models were proposed by Yu et al.,⁸ namely, the triangular and crosswise models. In this article, we focus our work on the NRRTM since it belongs to the admissible design and is more efficient. Under the NRRTM, an innocuous question with known population prevalence is introduced to indirectly obtain the answer to the sensitive question. Specifically, the question and answer options are placed in a 2×2 contingency table where two quadrants relate to the innocuous questions are with known population prevalence while the other two quadrants represent the binomial response options to the sensitive question. In the NRRTM, respondents are required to answer whether they belong to the No–No quadrant or any of the other three quadrants (i.e. Yes–No, Yes–Yes, or No–Yes). In this case, people admitting to the sensitive behavior are protected by the true answers of those who do not have the sensitive behavior to declare. Table 1 reports the survey design of the NRRTM. Let P be the random variable associated with the innocuous question (e.g. the last digit of your cell phone number is an even number), which is dichotomous and independent of the random variable Z associated with the sensitive question (e.g. ever had premarital sexual experience). Here, $Z=1$ represents the respondent possesses the sensitive characteristic; $=0$ otherwise, and $P=1$ represents the respondent possesses the non-sensitive characteristic; $=0$ otherwise. Let Y represent the answer in

Table 1. Traditional NRRT design for sensitive question.

Categories	$P=0$	$P=1$	Probabilities	$P=0$	$P=1$	Total
$Z=0$	○	•	$Z=0$	$(1-\pi)(1-p)$	$(1-\pi)p$	$1-\pi$
$Z=1$	•	•	$Z=1$	$\pi(1-p)$	πp	π
			Total	$(1-p)$	p	1

Table 2. The corresponding cell probabilities in the first group under DNRRTM (and ANRRTM).

Categories	$P=0$	$P=1$	Probabilities	$P=0$	$P=1$
$Z=0$	○	•	$Z=0$	$(1-\pi)(1-p_1) + \pi\theta$	
$Z=1$	•	•	$Z=1$	$\pi + (1-\pi)p_1 - \pi\theta$	

DNRRTM, dual non-randomized response triangular model; ANRRTM, alternating non-randomized response triangular model.

Table 3. The corresponding cell probabilities in the second group under DNRRTM.

Categories	$Q=0$	$Q=1$	Probabilities	$Q=0$	$Q=1$
$Z=0$	○	•	$Z=0$	$(1-\pi)(1-p_2) + \pi\theta$	
$Z=1$	•	•	$Z=1$	$\pi + (1-\pi)p_2 - \pi\theta$	

DNRRTM, dual non-randomized response triangular model.

the design: $Y=1$ means a tick is put in the triangular area and $Y=0$ means a tick is put in the circle. Under the NRRTM, $p = \Pr(P=1)$ is assumed to be known (e.g. $P=1$ represents the last digit of the respondent's cell number to be even and $p=0.5$) and $\pi = \Pr(Z=1)$ is the parameter to be estimated.

2.2 The dual non-randomized response triangular model (DNRRTM)

However, realizing that $Y=0$ means non-possession of the sensitive characteristic; some respondents with the sensitive characteristic (i.e. $Z=1$) will tend to provide the false answer (i.e. $Y=0$) due to guilty conscience. In order to take this kind of noncompliance into account, we consider the so-called dual non-randomized response triangular model with noncompliance (DNRRTM) in which two non-sensitive questions which are independent with the sensitive question are introduced. Here, we assume that the probability that a respondent belonging to the sensitive group will choose the safe answer (i.e. $Y=0$) to demonstrate his/her positive image due to guilty conscience with probability θ . Under DNRRTM, all respondents are randomly assigned to one of the following two groups. Respondents in the first (or second) group are asked to provide their answers to Table 2 (or Table 3) which combines the sensitive question and the first (second) non-sensitive question. That is, in both tables, the sensitive question is the same but the non-sensitive questions are different.

Similar to the NRRTM, $\pi = \Pr(Z=1)$ is the unknown parameter of interest. In the first (second) group, let P (Q) be the answer to the innocuous question with $P=1$ ($Q=1$) representing the

respondent possessing the characteristic of the non-sensitive question; and =0 otherwise. Here, $p_1 = \Pr(P = 1)$, $p_2 = \Pr(Q = 1)$ and $p_1 \neq p_2$ are assumed to be known or easily estimated.

Suppose there are r_1 (r_2) among n_1 (n_2) respondents who put a circle on the triangle formed by the three solid dots in the first (second) group. The likelihood function based on the data becomes

$$L(\pi, \theta | n_1, n_2, r_1, r_2, p_1, p_2) \propto [\pi + (1 - \pi)p_1 - \pi\theta]^{r_1} [(1 - \pi)(1 - p_1) + \pi\theta]^{n_1 - r_1} \\ \times [\pi + (1 - \pi)p_2 - \pi\theta]^{r_2} [(1 - \pi)(1 - p_2) + \pi\theta]^{n_2 - r_2}$$

The moment estimates (MEs) of π and θ can be easily shown to be

$$\begin{cases} \hat{\pi} = 1 + \frac{r_2/n_2 - r_1/n_1}{p_1 - p_2}, & \text{and} \\ \hat{\theta} = \frac{r_2(1 - p_1)/n_2 - r_1(1 - p_2)/n_1 + p_1 - p_2}{r_2/n_2 - r_1/n_1 + p_1 - p_2} \end{cases} \quad (1)$$

The asymptotic results for $\hat{\pi}$ and $\hat{\theta}$ are reported in the following theorem.

Theorem 1. *Let $\hat{\pi}$ and $\hat{\theta}$ be the moment estimates for π and θ given in equation (1). We have*

- (a) $E(\hat{\pi}) = \pi$
- (b) $\lim_{n_1, n_2 \rightarrow +\infty} E(\hat{\theta}) = \theta$
- (c) $\lim_{n_1, n_2 \rightarrow +\infty} \text{Var}(\hat{\pi}) \approx \frac{\lambda_1(1-\lambda_1) + \lambda_2(1-\lambda_2)\rho}{(p_1 - p_2)^2}$, where $\lambda_i = \pi + (1 - \pi)p_i - \pi\theta$, $i = 1, 2$, $\rho = \frac{n_1}{n_2}$ and
- (d) $\lim_{n_1, n_2 \rightarrow +\infty} \text{Var}(\hat{\theta}) \approx \frac{\lambda_1(1-\lambda_1)(1-p_2-\theta)^2}{\pi^2(p_1-p_2)^2} + \frac{\lambda_2(1-\lambda_2)(1-p_1-\theta)^2\rho}{\pi^2(p_1-p_2)^2}$

Theorem 1 suggests that $\hat{\pi}$ is an unbiased estimate of π while $\hat{\theta}$ is asymptotically unbiased estimate of θ . It should be noted that the point estimates $\hat{\pi}$ and $\hat{\theta}$ may sometimes fall outside the interval $[0, 1]$, which is meaningless in practice. In this case, the original NRRTM suggests truncating the estimate at zero (or one) which will always underestimate (or overestimate) the true proportion. For this issue, we consider the expectation-maximization algorithm to obtain a reliable estimate within $[0, 1]$. Here, let the number of respondents who are supposed to tick the triangle but eventually tick the circle be V_i in the i^{th} group, and W_i be the number of people whose answers to the sensitive question is “0” and to the non-sensitive question is “1”, $i = 1, 2$. Denote the missing data by $Y_{\text{mis}} = \{V_1, V_2, W_1, W_2\}$, and the complete data by $Y_{\text{com}} = \{Y_{\text{obs}}, Y_{\text{mis}}\}$ where $Y_{\text{obs}} = \{r_1, r_2\}$. Hence, the complete data likelihood function is given by

$$L(\pi, \theta | Y_{\text{com}}) = \pi^C (1 - \pi)^{n_1 + n_2 - C} \theta^{V_1 + V_2} (1 - \theta)^{r_1 + r_2 - W_1 - W_2}$$

where

$$C = r_1 + r_2 + V_1 + V_2 - W_1 - W_2$$

Hence, the M-step finds the maximum likelihood estimates of π and θ based on the complete data:

$$\pi = \frac{r_1 + r_2 + V_1 + V_2 - W_1 - W_2}{n_1 + n_2}, \quad \text{and} \quad \theta = \frac{V_1 + V_2}{r_1 + r_2 + V_1 + V_2 - W_1 - W_2}$$

The conditional predictive distributions are

$$V_i \sim \text{Binomial} \left(n_i - r_i, \frac{\pi\theta}{(1-\pi)(1-p_i) + \pi\theta} \right)$$

and

$$W_i \sim \text{Binomial} \left(r_i, \frac{(1-\pi)p_i}{\pi + (1-\pi)p_i - \pi\theta} \right) \quad i = 1, 2$$

while the E step replaces V_1, V_2, W_1, W_2 by their conditional expectations which are given by

$$V_i = (n_i - r_i) \times \frac{\pi\theta}{(1-\pi)(1-p_i) + \pi\theta}$$

and

$$W_i = r_i \times \frac{(1-\pi)p_i}{\pi + (1-\pi)p_i - \pi\theta} \quad i = 1, 2$$

It is noteworthy that using the expectation–maximization algorithm, the estimate always falls in the interval $[0, 1]$.

To our best knowledge, existing papers only proposed Wald-type confidence intervals which in practice yield lower bound less than zero or upper bound greater than one if the true value of π is close to zero or one. Furthermore, the Wald-type confidence interval constructions are based on normal approximation, which is questionable when the sample size is small to moderate. The bootstrap approach here can be used to obtain reliable confidence interval for π while the resultant confidence interval is guaranteed to lie within the interval $[0, 1]$. Using $\hat{\pi}$ and $\hat{\theta}$ obtained from the original data set, we generate r_i following $\text{Binomial}(n_i, \hat{\pi} + (1 - \hat{\pi})p_i - \hat{\pi}\hat{\theta})$, $i = 1, 2$. For each (r_1, r_2) , a bootstrap replication $\hat{\pi}^*$ by calculating the maximum likelihood estimates is obtained. Repeating this process G times, we get G bootstrap replications $\{\hat{\pi}_g^*\}_{g=1}^G$, and the bootstrap confidence interval is constructed as follows

$$[\hat{\pi}_{L(D)}, \hat{\pi}_{U(D)}]$$

where $\hat{\pi}_{L(D)}$ and $\hat{\pi}_{U(D)}$ are the $100(\alpha/2)$ and $100(1 - \alpha/2)$ percentiles of $\{\hat{\pi}_g^*\}_{g=1}^G$, respectively.

2.3 The alternating non-randomized response triangular model

In the DNRRTM, two different innocuous questions with different probabilities (i.e. $p_1 \neq p_2$) are required in the two groups. In this section, we propose a so-called alternating non-randomized response triangular model (ANRRTM) in which only one innocuous question P is included and the two categories of the non-sensitive question are alternated in the two groups. The survey design is substantially simplified since we need one non-sensitive question only. Here $p = \Pr(P = 1) \neq 0.5$ is known before the survey. Under the ANRRTM, the table associated with the first group is identical to Table 1 while that associated with the second group is reported in Table 4.

Table 4. The corresponding cell probabilities in the second group under ANRRTM.

Categories	$P=1$	$P=0$	Probabilities	$P=1$	$P=0$
$Z=0$	○	•	$Z=0$	$(1-\pi)p + \pi\theta$	
$Z=1$	•	•	$Z=1$	$1 - (1-\pi)p - \pi\theta$	

ANRRTM, alternating non-randomized response triangular model.

Suppose there are r_1 (r_2) among n_1 (n_2) respondents who put a circle in the triangle in the first (second) group. The likelihood based on the observed data is given by

$$L(\pi, \theta | n_1, n_2, r_1, r_2, p) \propto [\pi + (1-\pi)p - \pi\theta]^{r_1} [(1-\pi)(1-p) + \pi\theta]^{n_1-r_1} \\ \times [1 - (1-\pi)p - \pi\theta]^{r_2} [(1-\pi)p + \pi\theta]^{n_2-r_2}$$

The MEs for π and θ can be easily shown to be

$$\begin{cases} \hat{\pi} = \frac{r_2/n_2 - r_1/n_1}{2p-1} + 1, & \text{and} \\ \hat{\theta} = \frac{r_2(1-p)/n_2 - r_1p/n_1 + 2p-1}{r_2/n_2 - r_1/n_1 + 2p-1} \end{cases} \quad (2)$$

We have the following results for $\hat{\pi}$ and $\hat{\theta}$.

Theorem 2. Let $\hat{\pi}$ and $\hat{\theta}$ be the maximum likelihood estimates for π and θ given in equation (2). We have

- (a) $E(\hat{\pi}) = \pi$
- (b) $\lim_{n_1, n_2 \rightarrow +\infty} E(\hat{\theta}) = \theta$
- (c) $\lim_{n_1, n_2 \rightarrow +\infty} \text{Var}(\sqrt{n_1}\hat{\pi}) \approx \frac{\lambda'_1(1-\lambda'_1) + \lambda'_2(1-\lambda'_2)\rho}{(2p-1)^2}$, where $\lambda'_1 = \pi + (1-\pi)p - \pi\theta$, $\lambda'_2 = 1 - (1-\pi)p - \pi\theta$, $\rho = \frac{n_1}{n_2}$ and
- (d) $\lim_{n_1, n_2 \rightarrow +\infty} \text{Var}(\sqrt{n_1}\hat{\theta}) \approx \frac{\lambda'_1(1-\lambda'_1)(p-\theta)^2}{\pi^2(2p-1)^2} + \frac{\lambda'_2(1-\lambda'_2)(1-p-\theta)^2\rho}{\pi^2(2p-1)^2}$

Similarly, when π or θ are not in the interval $[0, 1]$ we employ the EM algorithm to obtain reliable estimates. Let the number of respondents who are supposed to tick triangle but eventually tick circle be V_i in the i^{th} group, and W_i be the number of people whose answers to the sensitive question is 0 and to the non-sensitive question is 1, $i = 1, 2$. Denote the missing data by $Y_{\text{mis}} = \{V_1, V_2, W_1, W_2\}$, and the complete data by $Y_{\text{com}} = \{Y_{\text{obs}}, Y_{\text{mis}}\}$, where $Y_{\text{obs}} = (r_1, r_2)$. Hence, the complete data likelihood function is given by

$$L(\pi, \theta | Y_{\text{com}}) = \pi^C (1-\pi)^{n_1+n_2-C} \theta^{V_1+V_2} (1-\theta)^{r_1+r_2-W_1-W_2}$$

where $C = r_1 + r_2 + V_1 + V_2 - W_1 - W_2$.

The M-step finds the maximum likelihood estimates of π and θ based on the complete data, i.e.

$$\pi = \frac{r_1 + r_2 + V_1 + V_2 - W_1 - W_2}{n_1 + n_2} \quad \text{and} \quad \theta = \frac{V_1 + V_2}{r_1 + r_2 + V_1 + V_2 - W_1 - W_2}$$

while the E-step replaces Z_1, Z_2, W_1, W_2 by its conditional expectations which are given by

$$V_1 = (n_1 - r_1) \times \frac{\pi\theta}{(1-\pi)(1-p) + \pi\theta}, \quad W_1 = r_1 \times \frac{(1-\pi)p}{\pi + (1-\pi)p - \pi\theta}$$

and

$$V_2 = (n_2 - r_2) \times \frac{\pi\theta}{(1-\pi)p + \pi\theta}, \quad W_2 = r_2 \times \frac{(1-\pi)(1-p)}{\pi + (1-\pi)(1-p) - \pi\theta}$$

Again, the expectation–maximization algorithm guarantees the estimates to fall in the interval $[0, 1]$.

Alternatively, we apply the bootstrap approach for confidence interval construction to guarantee the CI to fall into the interval $[0, 1]$. Using the obtained $\hat{\pi}$ and $\hat{\theta}$ from the original data, we generate $r_1 \sim \text{Binomial}(n_1, \hat{\pi} + (1 - \hat{\pi})p - \hat{\pi}\hat{\theta})$ and $r_2 \sim \text{Binomial}(n_2, \hat{\pi} + (1 - \hat{\pi})(1 - p) - \hat{\pi}\hat{\theta})$. For each (r_1, r_2) , we obtain a bootstrap replication $\hat{\pi}^*$. Repeating this process G times independently, we get G bootstrap replications $\{\hat{\pi}_g^*\}_{g=1}^G$. A $(1 - \alpha)100\%$ bootstrap CI for π can be calculated by

$$[\hat{\pi}_{L(A)}, \hat{\pi}_{U(A)}]$$

where $\hat{\pi}_{L(A)}$ and $\hat{\pi}_{U(A)}$ are the $100(\alpha/2)$ and $(1001 - \alpha/2)$ percentiles of $\{\hat{\pi}_g^*\}_{g=1}^G$, respectively.

3 Sample size determination

Sample size determination has become an important topic in surveys with sensitive questions. Accurate sample size formula is necessary since validity of statistical inferences from research studies heavily relies on this. Besides, sample size is important for both economic and ethical reasons. To test whether the population proportion is identical to a pre-specified value π_0 , we consider the following hypotheses

$$H_0 : \pi = \pi_0 \quad \text{versus} \quad H_1 : \pi \neq \pi_0$$

Under the null hypothesis, we have

$$\frac{\hat{\pi} - \pi_0}{\sqrt{\text{var}(\hat{\pi})}} \sim N(0, 1) \quad \text{as } n_1, n_2 \rightarrow +\infty$$

We first consider the DNRRTM. The null hypothesis is rejected at α level of significance if we observe the following

$$\hat{\pi} > \pi_0 + \frac{z_{1-\alpha/2}}{|p_1 - p_2|} \left\{ \frac{\lambda_1(1 - \lambda_1)}{n_1} + \frac{\lambda_2(1 - \lambda_2)}{n_2} \right\}^{1/2}$$

or

$$\hat{\pi} < \pi_0 - \frac{z_{1-\alpha/2}}{|p_1 - p_2|} \left\{ \frac{\lambda_1(1 - \lambda_1)}{n_1} + \frac{\lambda_2(1 - \lambda_2)}{n_2} \right\}^{1/2}$$

where $z_{1-\alpha/2}$ is the upper $\alpha/2$ quantile of the standard normal variable. Assume that $\pi = \pi_1$ and the power of the test can be approximated by

$$\begin{aligned} \text{Power(at } \pi_1) &= \alpha \text{ (rejecting } H_0 | \pi = \pi_1) \\ &= \Phi \left\{ \frac{|(\pi_0 - \pi_1)(p_1 - p_2)|n_1^{1/2}}{\sqrt{\lambda_3(1 - \lambda_3) + \rho\lambda_4(1 - \lambda_4)}} - Z_{1-\alpha/2} \left(\frac{\lambda_1(1 - \lambda_1) + \rho\lambda_2(1 - \lambda_2)}{\lambda_3(1 - \lambda_3) + \rho\lambda_4(1 - \lambda_4)} \right)^{1/2} \right\} \\ &= 1 - \beta \end{aligned}$$

where $\lambda_1 = \pi_0 + (1 - \pi_0)p_1 - \pi_0\theta$, $\lambda_2 = \pi_0 + (1 - \pi_0)p_2 - \pi_0\theta$, $\lambda_3 = \pi_1 + (1 - \pi_1)p_1 - \pi_1\theta$ and $\lambda_4 = \pi_1 + (1 - \pi_1)p_2 - \pi_1\theta$, $n_1 = \rho n_2$. Therefore, the sample sizes are given by

$$\begin{cases} n_1 = \frac{\left[Z_{1-\beta} \{ \lambda_3(1 - \lambda_3) + \rho\lambda_4(1 - \lambda_4) \}^{1/2} + Z_{1-\alpha/2} \{ \lambda_1(1 - \lambda_1) + \rho\lambda_2(1 - \lambda_2) \}^{1/2} \right]^2}{(\pi_0 - \pi_1)^2(p_1 - p_2)^2} \text{ and} \\ n_2 = n_1/\rho. \end{cases}$$

Similarly, the sample sizes based on ANRRTM can be obtained by

$$\begin{cases} n_1 = \frac{\left[Z_{1-\beta} \{ \lambda'_3(1 - \lambda'_3) + \rho\lambda'_4(1 - \lambda'_4) \}^{1/2} + Z_{1-\alpha/2} \{ \lambda'_1(1 - \lambda'_1) + \rho\lambda'_2(1 - \lambda'_2) \}^{1/2} \right]^2}{(\pi_0 - \pi_1)^2(2p - 1)^2} \text{ and} \\ n_2 = n_1/\rho \end{cases}$$

where $\lambda'_1 = \pi_0 + (1 - \pi_0)p - \pi_0\theta$, $\lambda'_2 = \pi_0 + (1 - \pi_0)(1 - p) - \pi_0\theta$, $\lambda'_3 = \pi_1 + (1 - \pi_1)p - \pi_1\theta$, and $\lambda'_4 = \pi_1 + (1 - \pi_1)(1 - p) - \pi_1\theta$, $n_1 = \rho n_2$

4 Simulation studies

In Section 2, we propose two NRRTs (i.e. DNRRTM and ANRRTM) which take the noncompliance into consideration. To investigate their performance, we consider biases of their estimates, the confidence widths and coverage probabilities.

In our simulation studies, we consider $\pi = 0.05, 0.1, 0.15, 0.2, 0.3, 0.4$, $\theta = 0.1, 0.25$, and $n_1 = n_2 = 2000$. For the DNRRTM, consider two settings for the proportions of the two innocuous questions: $(p_1, p_2) = (0.75, 0.3)$ for $\theta = 0.1$ and $(p_1, p_2) = (0.7, 0.2)$ for $\theta = 0.25$. For ANRRTM, we consider similar settings for the proportion of the single innocuous question: $p = 0.3$ for $\theta = 0.1$ and $p = 0.25$ for $\theta = 0.25$. Here, r_1 follows Binomial $(n_1, \pi + (1 - \pi)p_1 - \pi\theta)$ and r_2 follows Binomial $(n_2, \pi + (1 - \pi)p_2 - \pi\theta)$ under DNRRTM while r_1 follows Binomial $(n_1, \pi + (1 - \pi)p - \pi\theta)$ and r_2 follows Binomial $(n_2, \pi + (1 - \pi)(1 - p) - \pi\theta)$ under ANRRTM. For each possible configuration, we simulate 1000 pairs of (r_1, r_2) and hence obtain 1000 estimates and bootstrap confidence intervals of π . We report the mean of the 1000 estimates (denoted as

Table 5. Estimates and confidence intervals using NRRT and DNRRTM for $\theta = 0.10$, $p_1 = 0.75$, $p_2 = 0.3$.

	NRRT		DNRRTM	
	$\hat{\pi}$	Width of CI	$\hat{\pi}$	Width of CI
$\pi = 0.05$	0.038	0.0879 (93.0%)	0.066	0.079 (80.5%)
$\pi = 0.10$	0.073	0.0878 (73.5%)	0.110	0.082 (88.5%)
$\pi = 0.15$	0.109	0.0875 (55.1%)	0.157	0.086 (92.9%)
$\pi = 0.20$	0.146	0.0871 (37.0%)	0.202	0.090 (93.4%)
$\pi = 0.30$	0.219	0.0859 (9.4%)	0.301	0.096 (90.2%)
$\pi = 0.40$	0.292	0.0841 (1.5%)	0.400	0.100 (89.2%)

NRRT, non-randomized response technique; DNRRTM, dual non-randomized response triangular model.

Table 6. Estimates and confidence intervals using NRRT and DNRRTM for $\theta = 0.25$, $p_1 = 0.7$, $p_2 = 0.2$.

	NRRT		DNRRTM	
	$\hat{\pi}$	Width of CI	$\hat{\pi}$	Width of CI
$\pi = 0.05$	0.026	0.0754 (81.0%)	0.058	0.071 (91.4%)
$\pi = 0.10$	0.045	0.0761 (18.9%)	0.102	0.079 (94.6%)
$\pi = 0.15$	0.066	0.0766 (2.5%)	0.151	0.088 (90.6%)
$\pi = 0.20$	0.088	0.0771 (0.0%)	0.198	0.093 (90.3%)
$\pi = 0.30$	0.128	0.0776 (0.0%)	0.300	0.097 (88.9%)
$\pi = 0.40$	0.170	0.0780 (0.0%)	0.401	0.098 (91.9%)

NRRT, non-randomized response technique; DNRRTM, dual non-randomized response triangular model.

Table 7. Estimates and confidence intervals using NRRT and ANRRTM for $\theta = 0.10$, $p = 0.25$.

	NRRT		ANRRTM	
	$\hat{\pi}$	Width of CI	$\hat{\pi}$	Width of CI
$\pi = 0.05$	0.039	0.0853 (92.0%)	0.063	0.070 (80.0%)
$\pi = 0.10$	0.073	0.0854 (73.6%)	0.109	0.074 (89.5%)
$\pi = 0.15$	0.110	0.0853 (55.7%)	0.154	0.077 (93.2%)
$\pi = 0.20$	0.147	0.0850 (36.4%)	0.203	0.081 (95.0%)
$\pi = 0.30$	0.219	0.0840 (9.4%)	0.304	0.088 (91.1%)
$\pi = 0.40$	0.294	0.0825 (1.2%)	0.402	0.092 (90.7%)

NRRT, non-randomized response technique; ANRRTM, alternating non-randomized response triangular model.

$\hat{\pi}$), and the mean width of the 1000 confidence intervals with significance level 0.1 and its coverage probability in Tables 5 and 6 for DNRRTM. Similar results are reported in Tables 7 and 8 for ANRRTM. To show the effect of noncompliance, we also include the results based on the original NRRT for comparison purpose. In NRRT, the estimate is

Table 8. Estimates and confidence intervals using NRRT and ANRRTM for $\theta = 0.25$, $p = 0.2$

	NRRT		ANRRTM	
	$\hat{\pi}$	Width of CI	$\hat{\pi}$	Width of CI
$\pi = 0.05$	0.024	0.0928 (80.9%)	0.054	0.059 (92.1%)
$\pi = 0.10$	0.038	0.0936 (14.9%)	0.099	0.067 (91.2%)
$\pi = 0.15$	0.054	0.0944 (1.1%)	0.150	0.073 (90.6%)
$\pi = 0.20$	0.070	0.0950 (0.0%)	0.200	0.076 (89.1%)
$\pi = 0.30$	0.104	0.0960 (0.0%)	0.301	0.078 (89.0%)
$\pi = 0.40$	0.137	0.0965 (0.0%)	0.400	0.079 (89.6%)

NRRT, non-randomized response technique; ANRRTM, alternating non-randomized response triangular model.

obtained by taking the mean of the point estimates in two groups based on the original NRRT method.

According to Tables 5 and 7, NRRT yields severely biased estimates and extremely low coverage probabilities even for small proportion of noncompliance (i.e. $\theta = 0.1$). On the other hand, the simulation results based on DNRRTM and ANRRTM are satisfactory as it can provide more reliable estimates and confidence intervals. As shown in Tables 6 and 8, when the proportion of cheating increases to 0.25 (which is adopted by the experiment studies by Edgell et al.¹²), the performance of NRRT is even worse. In particular, its bias is nearly half of the true value and its coverage probability can be as low as 0 for $\pi = 0.3$ or 0.4. For DNRRTM and ANRRTM, we observe that when π increases, the bias becomes smaller and the confidence width becomes wider. It should be noted that direct comparison between DNRRTM and ANRRTM is not possible as the former involves two innocuous questions. Based on our simulation results, both ANRRTM and DNRRTM are safely and strongly recommended in survey designs with sensitive questions. In terms of simplicity, the ANRRTM is recommended.

5 Real example: pre-marital sex experience

The risk of transmission of HIV and other sexually transmitted diseases is higher in sexual relationships with multiple partners and without the use of condoms. Premarital sex often involves multiple partners, and extramarital sex, by definition, implies multi-partner relationships. Avoidance of multi-partner sexual relationships, use of condoms and sexual abstinence are usually advocated for prevention of spread of HIV and other sexually transmitted diseases. In some Asian countries, premarital sexual activity has long been considered as taboo and becomes a sensitive issue in which we can hardly get reliable answer by direct asking. To investigate the proportion of college students (in Changchun City of Jilin) who have ever had premarital sex experience, we apply the proposed DNRRTM and ANRRTM to estimate the desired proportion. In both surveys, the sensitive variable (i.e. Y) represents whether an interviewee has had premarital sex intercourse (i.e. $Y = 1$ if yes; $= 0$ otherwise). Under the DNRRTM, the non-sensitive variable (i.e. P) in the first group represents whether the last digit of an interviewee's phone number is odd (i.e. $P = 1$ if yes; $= 0$ otherwise) while in the second group the non-sensitive variable (i.e. Q) represents whether the interviewee was born in the first nine months of a year (i.e. $Q = 1$ if yes; $= 0$ otherwise). At the end of the study, we observed $n_1 = 97$, $r_1 = 60$, $n_2 = 76$, $r_2 = 60$, $p_1 = 0.5$, and $p_2 = 0.75$. Simple calculations

yield $\hat{\pi} = 0.3163$ and the 90% confidence interval for π is equal to (0.1387, 0.7145). If we assume that all respondents provide their truthful answers according to the survey design and apply the NRRT, the estimate for π is $\hat{\pi} = 0.1975$ which is substantially smaller than the estimate based on DNRRTM.

Under the ANRRTM, the non-sensitive variable (i.e. P) in the first group represents whether the interviewee was born in the first nine months of a year (i.e. $P = 1$ if yes; $=0$ otherwise). At the end of the study, we observed $n_1 = 76$, $r_1 = 60$, $p_1 = 0.75$, and $n_2 = 78$, $r_2 = 33$, and $p_2 = 0.25$. Simple calculations yield $\hat{\pi} = 0.2672$ with 90% confidence interval for π being (0.1404, 0.5007). Using the NRRT, the estimate is $\hat{\pi} = 0.1943$. The difference of the two estimates strongly suggests that the noncompliance indeed exists in the survey.

6 Conclusion

In this article, we consider two design techniques which incorporate the cheating behavior (i.e. noncompliance) into the NRRTM. They are namely the DNRRTM and ANRRTM. The fundamental difference between the two noncompliance non-randomized response models is the number of innocuous questions being introduced in the design. According to our simulation results, both proposed models are superior to the existing NRRTM in the sense that they generally yield unbiased estimates and guarantee the coverage probabilities close to the pre-specified coverage level. On the other hand, the existing NRRTM could produce severely biased estimates and confidence intervals with zero coverage probability. In practice, the ANRRTM is highly recommended as it introduces only one innocuous question in the questionnaire.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of interest

None declared.

References

1. Warner SL. Randomized response: a survey technique for eliminating evasive answer bias. *J Am Stat Assoc* 1965; **60**: 63–69.
2. Miller JD. *A new survey technique for studying deviant behavior*. PhD Thesis, George Washington University, New York, USA, 1984.
3. Tian GL, Yu JW, Tang ML, et al. A new non-randomized model for analyzing sensitive questions with binary outcomes. *Stat Med* 2007; **26**: 4238–4252.
4. Greenberg BG, Kuebler RR, Abernathy JR, et al. Application of the randomized response technique in obtaining quantitative data. *J Am Stat Assoc* 1971; **66**: 243–250.
5. Kuk AYC, Abernathy JR and Horvitz DG. Asking sensitive questions indirectly. *Biometrika* 1990; **77**: 436–438.
6. Mangat NS and Singh R. An alternative randomized response procedure. *Biometrika* 1990; **77**: 439–442.
7. Böckenholt U and Van der Heijden PGM. Item randomized-response models for measuring noncompliance: risk-return perceptions, social influences, and self-protective responses. *Psychometrika* 2007; **72**: 245–262.
8. Yu JW, Tian GL and Tang ML. Two new models for survey sampling with sensitive characteristic: design and analysis. *Metrika* 2008; **67**: 251–263.
9. Nayak TK. On randomized response surveys for estimating a proportion. *Commun Stat – Theory Methods* 1994; **23**: 3303–3321.
10. Petróczi A, Nepusz T, Cross P, et al. New non-randomised model to assess the prevalence of discriminating behaviour: a pilot study on mephedrone. *Substance Abuse Treat, Prevent, Policy* 2011; **6**: 1–20.
11. Peng X and Yan ZZ. Two-valued response technique of polychotomous sensitive question. *J Syst Sci Complex* 2011; **6**: 1193–1203.

12. Edgell SE, Himmelfarb S and Duchan KL. Validity of forced responses in a randomized response model. *Soc Methods Res* 1982; **11**: 89–100.
13. van der Heijden PGM, Gils GV, Bouts J, et al. A comparison of randomized response, computer-assisted self-interview, and face-to-face direct questioning eliciting sensitive information in the context of welfare and unemployment benefit. *Soc Methods Res* 2000; **28**: 505–537.
14. Zhou H, Wang XY, Ye F, et al. Contraceptive knowledge, attitudes and behavior about sexuality among college students in Beijing, China. *Chinese Med J* 2012; **125**: 1153–1157.
15. He L. Premarital sex among never married young adults in contemporary China: comparison between males and females. In: *At the population association of American 2012 Annual meeting program in San Francisco*. San Francisco, CA, 3–5 May 2012.
16. Cruyff Maarten JLF, van den Hout A, van der Heijden PGM, et al. Log-linear randomized-response models taking self-protective response behavior into account. *Soc Methods Res* 2007; **3**: 266–282.
17. Böckenholt U, Barlas S and Van Der Heijden PGM. Do randomized-response designs eliminate response biases? An empirical study of non-compliance behavior. *J Appl Econometr* 2009; **24**: 377–392.