Received: 9 May 2019 | Revised: 10 October 2019 | Accepted: 31 October 2019
DOI: 10.1002/cmm4.1073

REVIEW ARTICLE

WILEY

Variance jackknife estimation for randomized response surveys: A simulation study and an application to explore cheating in exams and bullying

Raghunath Arnab[1] | **Beatriz Cobo**[2]

[1]Department of Statistics, University of Botswana, Gaborone, Botswana

[2]Department of Statistics and Operational Research, Faculty of Mathematical Sciences, Complutense University of Madrid, Madrid, Spain

Correspondence
Beatriz Cobo, Department of Statistics and Operational Research, Faculty of Mathematical Sciences, Complutense University of Madrid, Plaza Ciencias, 3, 28040 Madrid, Spain.
Email: beacr@ugr.es

Funding information
Ministerio de Economía y Competitividad, Grant/Award Number: MTM2015-63609-R

The aim of the randomized response (RR) technique is to decrease social desirability bias, thus guaranteeing confidentiality, improving respondent cooperation, and obtaining reliable estimates. This technique obtains stronger estimates of sensitive characteristics, compared to direct questioning, by reducing respondents motivation to report falsely their attitudes. In this paper, we have considered the problem of estimation of variance of an estimator based on an RR survey data obtained from a complex survey design by using resampling techniques. The performance of the proposed jackknife modified estimator is tested by simulation studies. The simulation studies reveal that there is a decrease in the relative bias and the relative mean square error for all RR schemes. An RR survey was conducted to investigate sensitive behaviors such as cheating in exams and bullying. The application of the new jackknife confidence intervals revealed substantial increase in high-tech cheating and bullying in the university campuses.

KEYWORDS
complex surveys, randomized response, resampling methods

1 | INTRODUCTION

In socioeconomic and biometric research, we very often gather information relating to highly sensitive issues such as induced abortion, drug addiction, HIV infection status, duration of suffering from AIDS, sexual behavior, incidence of domestic violence, tax evasion, and so on. In these situations using the direct method of interview (asking questions directly to the respondents), the respondents provide often untrue response or even refuse to respond because of the social stigma and or fear.

Survey statisticians and practitioners have developed many different strategies to ensure interviewees' anonymity and to reduce the incidence of evasive answers and underreporting of social taboos when direct questions are posed on sensitive issues. One possibility is to limit the influence of the interviewer, by providing self-administered questionnaires (SAQs) with paper and pencil, the interactive voice response technique, computer-assisted telephone interviewing, computer-assisted self-interviewing (CASI), audio CASI, or by computer-assisted web interviewing. Alternatively, since the 1960s, a variety of questioning methods have been devised to ensure respondents' anonymity and to reduce the incidence of evasive answers and the over/underreporting of socially undesirable acts. These methods are generally known as indirect questioning techniques (IQTs) (for a review, see the work of Chaudhuri and Christofides,[1] and they obey the principle that no direct question is posed to survey participants. Therefore, there is no need for respondents to openly reveal whether they have actually engaged in activities or present attitudes that are socially sensitive. Their privacy is protected because the responses remain confidential to the respondents, and consequently, their true status remains uncertain and undisclosed to both the interviewer and the researcher. Nonetheless, although the individual information, provided by the respondents according to the rules prescribed by the adopted IQT, cannot be used to discover their true status regarding the sensitive issues, the information gathered for all the survey participants can be profitably employed to draw inferences on certain parameters of interest for the study population. The IQTs comprise various strategies for eliciting sensitive information, which mainly encompass these approaches: the randomized response (RR) technique (RRT), the item count technique, and the nonrandomized response technique. All the approaches have produced a considerable literature and attracted the interest of health, cognitive, and behavioral psychologists, epidemiologists, health care operators, researchers engaged in organizing, managing, and conduction sensitive studies, as well as policy makers committed in formulating effective diseases and mental disorders control measures and promoting public intervention programs to gauge progress toward improving the behavioral health of a state. In terms of the volume of research conducted in this field since Warner's[2] pioneering work on indirect questioning, the RRT maintains a prominent position among IQTs.

Contextually, many studies have assessed the validity of RR methods, showing that they can produce more reliable answers than conventional data collection methods (eg, direct questioning (DQ) in face-to face interviews, SAQs with paper and pencil and CASIs). The RRT has been applied in surveys covering a variety of sensitive topics, as drug use, racism, abortion, sexual victimization, cheating, and so on, as indicated in the articles of Perri et al[3] and Rueda et al.[4] We found some more recent studies related with people's sexual practice in the United States,[5] health care study of the dependence structure,[6] inappropriate sexual behavior among university students,[7] contaminants in the air and water,[8] genital cutting in Southern Ethiopia,[9] female sex workers in Taiyuan, China,[10], drug use in Republic of Georgia,[11] illegal cable TV subscribers,[12] and the right to receive welfare.[13]

Most of the surveys in practice are complex surveys, involving stratification, clustering, and unequal probability of selection of sample. In such surveys, information of more than one character is collected at a time. Some of them are confidential in nature while the others are not. In such a complex survey design, unbiased variance estimation is not easy because of clustering and involvement of second-order inclusion probabilities that are generally complex. Apart from these difficulties, variance estimators may not always be nonnegative. The variance estimator is important to find the magnitude of sampling error and determine the confidence intervals of the parameters of interest and optimal sample size for conducting a survey. Various methods of variance estimation from complex survey designs are listed by Wolter.[14] Popular among them are linearization, random group, jackknife (JK), balance repeated replication, and bootstrap. A little has been done to compare relative performances of these variance estimators. Limited empirical studies reveal that none of the proposed variance estimators is the best in all situations.[14]

In this paper, we consider the JK estimation of variance of an estimator based on a RR survey data obtained from a complex survey design. The following section gives a short description about the estimation when data are obtained by RR procedures. In Section 3, we show that the JK variance estimator underestimates the variance, and we propose a modification of the conventional JK variance estimator. Results of the simulation studies regarding the finite sample performance of proposed estimator are reported in Section 4. Section 5 shows an application of the proposed method for data from a real survey and, finally, Section 6 presents the conclusions drawn.

2 | MATERIALS AND METHODS

2.1 | Randomized response techniques

Consider a finite population $U = \{1, \ldots, i, \ldots, N\}$, consisting of N different elements. Let y_i be the value of the sensitive aspect under study for the ith population element. Our aim is to estimate the finite population total $Y = \sum_{i=1}^{N} y_i$ or the population mean $\overline{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i$. In case the study variable y_i is qualitative, we define $y_i = 1$ if the ith unit of the population possesses certain attribute A (say) and $y_i = 0$ if the unit does not possess the attribute A. In this case, the population mean $\overline{Y}$ is equal to the population proportion π of population units bearing the sensitive attribute A.

Assume that a sample s of individuals is chosen according to a general design $p(\cdot)$, which admits positive first-order and second-order inclusion probabilities $\pi_i = \sum_{i \in s} p(s)$ and $\pi_{ij} = \sum_{i,j \in s} p(s)$ with $i, j \in U$. For the sake of notation, let $d_i = \pi_i^{-1}, d_{ij} = \pi_{ij}^{-1}, \Delta_{ij} = \pi_{ij} - \pi_i \pi_j$. Under a DQ survey mode, let $\hat{\overline{Y}}_{\text{HT}}$ denote the well-known Horvitz-Thompson estimator (hereafter HT-estimator[15]) of $\overline{Y}$:

$$\hat{\overline{Y}}_{\text{HT}} = \frac{1}{N} \sum_{i \in s} d_i y_i.$$

The estimator is unbiased and has variance

$$V\left(\hat{\bar{Y}}_{\mathrm{HT}}\right)=\frac{1}{N^2}\sum_{i\in U}\sum_{j\in U}\Delta_{ij}d_iy_id_jy_j,$$

which can be unbiasedly estimated by

$$\hat{V}\left(\hat{\bar{Y}}_{\mathrm{HT}}\right)=\frac{1}{N^2}\sum_{i\in s}\sum_{j\in s}d_{ij}\Delta_{ij}d_iy_id_jy_j.$$

To consider a wide variety of RR procedures, we consider the unified approach given by Arnab.[16] The interviews of individuals in the sample s are conducted in accordance with an RR model. The variable y is the sensitive variable, and we cannot observe it directly from the respondent, so it will be estimated through the RR obtained from the ith respondent. Suppose that the ith respondent has to conduct a RR trial independently and z_i is the RR (or scrambled response) for the trial. For each $i\in s$, the RR induces a revised RR, r_i such as $E_R(r_i)=y_i$ and $V_R(r_i)=\phi_i$ where the operators E_R and V_R denote expectation and variance with respect to randomization procedure RR. As usual, in the design-based approach to RRTs, it is assumed that the sampling design and the randomization stage are independent of each other (eg, Barabesi et al[17]) and that the randomization stage is performed on each selected individual independently.

In this general setup, the HT-type estimator for the population mean of the sensitive characteristic y is given by

$$\hat{\bar{Y}}_{\mathrm{HT}}(r)=\frac{1}{N}\sum_{i\in s}d_ir_i.$$

The variance of $\hat{\bar{Y}}_{\mathrm{HT}}(r)$ (see the work of Chaudhuri et al[18]) is given by

$$V\left(\hat{\bar{Y}}_{\mathrm{HT}}(r)\right)=\frac{1}{N^2}\left(\sum_{i\in U}\sum_{j\in U}\Delta_{ij}d_iy_id_jy_j+\sum_{i\in U}d_iV_R(r_i)\right)=\frac{1}{N^2}\left(V\left(\hat{\bar{Y}}_{\mathrm{HT}}\right)+\sum_{i\in U}d_i\phi_i\right)$$

and an estimator of $V\,(\hat{\bar{Y}}_{\mathrm{HT}}(r))$ is

$$\begin{aligned}\hat{V}\left(\hat{\bar{Y}}_{\mathrm{HT}}(r)\right)&=\frac{1}{N^2}\left(\sum_{i\in s}\sum_{j\in s}d_{ij}\Delta_{ij}d_ir_id_jr_j+\sum_{i\in s}d_i^2\hat{V}_R(r_i)\right)\\&=\frac{1}{N^2}\left(\hat{V}\left(\hat{R}_{\mathrm{HT}}\right)+\sum_{i\in s}d_i^2\hat{\phi}_i\right),\end{aligned}$$

where $\hat{V}_R(r_i)$ is an unbiased estimator for $V_R(r_i)$ of the considered RR model.

This estimator is an unbiased estimator of $V(\hat{\bar{Y}}_{\mathrm{HT}}(r))$ if $\hat{V}(\hat{R}_{\mathrm{HT}})$ is an RR-unbiases for $V(\hat{\bar{Y}}_{\mathrm{HT}})$.

In qualitative models, the values r_i and $V_R(r_i)$ for $i\in s$ are described in each model.

For example, in the case of Warner model, a sampled person labeled i is offered a box of a considerable number of identical cards with a proportion p; $(0<p<1;\,p\neq 0.5)$ of them marked A, the senstive attribute, and the rest marked A^c. The person is requested, randomly, to draw one of them, to observe the mark on the card, and to give the response

$$z_i=\begin{cases}1,\text{if card type "matches" the trait A or A}^{\mathrm{c}}\\0,\text{if a "no match" results}\end{cases}$$

So, the expectation of the variable z is

$$E_R(z_i)=py_i+(1-p)(1-y_i)=(1-p)(2p-1)y_i.$$

The transformed variable is

$$r_i = \frac{z_i - (1-p)}{2p-1}$$

and $E_R(r_i) = y_i$.

Since $z_i^2 = z_i$ and $y_i^2 = y_i$

$$V_R(z_i) = E_R(z_i)\,(1 - E_R(z_i)) = p(1-p)$$

So

$$V_R(r_i) = \frac{p(1-p)}{(2p-1)^2}$$

Now, noting $V_R(r_i) = E_R(r_i^2) - (E_R(r_i))^2 = E_R(r_i^2) - y_i^2 = E_R(r_i^2) - y_i = E_R(r_i^2) - E_R(r_i) = E_R(r_i(r_i-1))$, we set a general unbiased estimator of $V_R(r_i)$ as

$$\hat{V}_R(r_i) = r_i(r_i - 1).$$

In some quantitative models, the values r_i and $V_R(r_i)$ for $i \in s$ are calculated in a general form in the article of Arcos et al.[19]

To define this estimator we need π_{ij} for all units in each frame and $\pi_{ij} > 0$ for all units. In some common sampling designs (as cluster sampling with probability proportional to size) these probabilities are unknown or can be equal to 0 for some sampling units i, j. A simple alternative is the use of with replacement variance estimators[20 (pp99)] or replicated sampling methods (see the work of Wolter[14] for a detailed description of these techniques in finite population sampling). Quenouille[21] introduced the JK method to estimate the bias of an estimator by deleting one datum each time from the original data set and recalculating the estimator based on the rest of the data. In survey sampling it is usual to use JK techniques due to their simplicity and because they are implemented in general purpose software packages, such as R.

Now, we explore the possibility of using JK methods to estimate the variance of the estimators in RRT.

3 | JACKKNIFE METHODS TO ESTIMATE THE VARIANCE OF THE ESTIMATORS IN RRT

3.1 | Nonstratified and non clustered sampling designs

For a nonstratified and nonclustered designs, the JK variance estimator of $V(\hat{\hat{Y}}_{\text{HT}}(r))$ is given by

$$\hat{V}_{\text{JK}}\left(\hat{\hat{Y}}_{\text{HT}}(r)\right) = \frac{n-1}{n}\sum_{i=1}^{n}\left(\hat{\hat{Y}}_{\text{HT}(-i)}(r) - \bar{\hat{\hat{Y}}}_{\text{HT}}(r)\right)^2,$$

where $\hat{\hat{Y}}_{\text{HT}(-i)}(r) = \frac{1}{N}\sum_{j\in s-i}\frac{r_j}{\left(\frac{n-1}{n}\pi_j\right)}$ is the value taken by $\hat{\hat{Y}}_{\text{HT}}(r)$ after dropping unit i from sample s and $\bar{\hat{\hat{Y}}}_{\text{HT}}(r) = \frac{1}{n}\sum_{i=1}^{n}\hat{\hat{Y}}_{\text{HT}(-i)}(r)$ is the mean of values $\hat{\hat{Y}}_{\text{HT}(-i)}(r)$.

We will show in the following Theorem that $\hat{V}_{\text{JK}}(\hat{\hat{Y}}_{\text{HT}}(r))$ underestimates the variance $V(\hat{\hat{Y}}_{\text{HT}}(r))$.

Theorem 1. *Let p(s) be a nonstratified and nonclustered design with π_i and π_{ij} be the first and second-order inclusion probabilities. We assume that survey is conducted in accordance with an RR model and r_i is an unbiased estimation of y_i with $V(r_i) = \phi_i$. Under these assumptions, the estimator $\hat{V}_{\text{JK}}(\hat{\hat{Y}}_{\text{HT}}(r))$ asymptotically underestimate the variance $V(\hat{\hat{Y}}_{\text{HT}}(r))$ and the amount of bias is $-\frac{1}{N^2}\sum_{i\in U\phi i}$*

Proof. See Supplementary Material. □

Theorem 2. *Under the same assumptions as Theorem 1, the adjusted variance estimator*

$$\hat{V}_{\text{JK}}^{ad}\left(\hat{\hat{Y}}_{\text{HT}}(r)\right) = \hat{V}_{\text{JK}}\left(\hat{\hat{Y}}_{\text{HT}}(r)\right) + \frac{1}{N^2}\sum_{i\in s} d_i\hat{\phi}_i$$

is asymptotically unbiased for $V(\hat{\hat{Y}}_{\text{HT}}(r))$ in the sense that $E(\hat{V}_{\text{JK}}^{ad}(\hat{\hat{Y}}_{\text{HT}}(r))) \cong V(\hat{\hat{Y}}_{\text{HT}}(r))$.

Proof. See Supplementary Material. □

3.2 | Stratified sampling design

Stratification is a common technique used in surveys to reduce variances. Stratified RR models are considered in Kim and Elam.[22] The JK runs into some difficulty in the context of stratified sampling plans because the observations are no longer identically distributed. We shall describes one methods for handling this problem.

We assume the population is divided into L strata, where N_h describes the size of the hth stratum. Sampling is carried out independently in the various strata. From stratum h, a sample s_h is selected by a sampling design with inclusion probabilities π_{hi} for the ith unit of stratum h and π_{hij} for the ith and jth units of the stratum h. We assume that, for each unit in the sample s_h, the RR induces a random variable z_{hi} so that the revised RR r_{hi} is an unbiased estimation of y_{hi} with $V(r_{hi}) = \phi_{hi}$. We define an adjusted stratified JK estimator as

$$\hat{V}^{adst}_{\text{JK}}\left(\hat{\overline{\text{Y}}}_{\text{HT}}(r)\right) = \sum_{h=1}^{L} \frac{n_h - 1}{n_h} \sum_{i=1}^{n_h} \left(\hat{\overline{\text{Y}}}_{\text{HT}(-hi)}(r) - \hat{\overline{Y}}_{\text{HTh}}(r)\right)^2 + \frac{1}{N^2} \sum_{h=1}^{L} \sum_{i \in s_h} \frac{\hat{\phi}_{hi}}{\pi_{hi}},$$

where $\hat{\overline{Y}}_{\text{HT}(-hi)}(r)$ is the value taken by $\hat{\overline{Y}}_{\text{HT}}(r)$ after dropping unit i of stratum h from sample s_h, $\hat{\overline{Y}}_{\text{Hth}}(r)$ is the mean of values $\hat{\overline{Y}}_{\text{HT}(-i)}(r)$ and $\hat{\phi}_{hi}$ is an RR-unbiased estimator of ϕ_{hi}.

Theorem 3. *Let p(s) be a stratified design. Under the above randomized device in each stratum sample s_h, the stratified estimator $\hat{V}^{adst}_{\text{JK}}(\hat{\overline{Y}}_{\text{HT}}(r))$ is asymptotically unbiased for the variance.*

Proof. See Supplementary Material. □

3.3 | Clustered sampling designs

A nonclustered sampling design is assumed previously. No new principles are involved in the application of JK methodology to clustered samples. We simple work with the ultimate cluster rather than elementary units.

In the following section, we will compare the performance of proposed adjusted JK estimator with alternative estimators respect to bias, proportion of times of underestimations and efficiency through simulation studies.

4 | SIMULATION STUDY

The simulation study is based on qualitative and quantitative models. Our simulations were programmed in R and new R code was developed to compute the estimators compared.

4.1 | Qualitative variables

In the first simulation study, we have considered three populations each of size $N = 1000$ generated by drawing samples from Bernoulli population with parameter $\pi = 0.05$, 0.15, and 0.25. We will denote them by Population I, Population II, and Population III respectively.

To estimate $\overline{Y}$ we use different RR models and for each case, 1000 samples of size $n = 500$ were drawn from the population using simple random sampling without replacement.

From each of the selected samples s from qualitative populations (Population I, Population II, and Population III) randomized data was generated by using three RR models R_t, $t = 1,2$ and 3. The first RR model (R_1) is the Warner device[2] with the parameter $p = 0.6$, R_2 parameter model is the Horvitz model (see the works of Horvitz et al[23] and Greenberg et al[24]) with the parameter $p = 0.6$, and the third model R_3 is the forced response model proposed by Boruch[25] with the parameters $p_1 = 0.6$,$p_2 = p_3 = 0.2$.

We denote $\hat{\theta} = \hat{V}(\hat{\overline{Y}}_{\text{HT}}(r)), \hat{V}_{\text{JK}}(\hat{\overline{Y}}_{\text{HT}}(r)), \hat{V}^{ad}_{\text{JK}}(\hat{\overline{Y}}_{\text{HT}}(r))$, based on the model R_t(for $t = 1,2,3$) by $\hat{\theta}_t$. V_t denotes the empirical variance of the estimator $\hat{\overline{Y}}_{\text{HT}}(r)$ for the model R_t. We denote

$$V_t = \frac{1}{(100000)^2} \sum_{i=1}^{100000} \left(\hat{\text{Y}}_{\text{HT}(i)}(r) - \hat{\overline{Y}}_{\text{HT}}(r)\right)^2,$$

where

$$\hat{\hat{\overline{Y}}}_{\mathrm{HT}}(r) = \frac{1}{100000}\sum_{i=1}^{100000}\hat{Y}_{\mathrm{HT}(i)}(r).$$

The selection of sample is repeated $Q(=1000)$ times. The based $\hat{\theta}_t$ on the $q(=1,\ldots,Q)$th repetition will be denoted by $\hat{\theta}_{t|q}$. The relative bias (RB) and the relative mean square errors (RMSEs) of the estimator $\hat{\theta}_t$ are calculated as follows:

$$\mathrm{RB}(\hat{\theta}_t) = \frac{1}{Q}\sum_{q=1}^{Q}\frac{\left(\hat{\theta}_{t|q} - V_t\right)}{V_t},$$

$$\mathrm{RMSE}(\hat{\theta}_t) = \frac{1}{Q}\sum_{q=1}^{Q}\frac{\left(\hat{\theta}_{t|q} - V_t\right)^2}{V_t^2}$$

We also compute the proportion of underestimation of variance (NB), ie, proportion of times bias becomes negative. For each sample q, we also calculate the JK confidence interval $(1-\alpha)\%$ level for the mean $\overline{Y}$ as $(\hat{\hat{\overline{Y}}}_{\mathrm{HT}}(r) \pm z_{\alpha/2}\sqrt{\hat{\theta}_t})$ where $z_{\alpha/2}$ denotes the $(1-\alpha)\%$ quantile of a standard normal distribution (we consider $\alpha = 0.05$). We calculate the real coverage (COVE) as the proportion of times that the parameter $\overline{Y}$ is within the interval.

We note that the given model parameters for the three RR models provide different levels of privacy protection and consequently different risks of nonresponse and untruthful answering. Hence, the resulting variances must not be compared.

To check behavior of estimators when applied to real data selected through complex sampling designs, we are going to consider data from a real population of 1500 families from a Spanish autonomous community, specifically Andalusia,[26,27] in which some characteristics are studied; among them, we consider as sensitive variable if the family pay the mortgage or not. The proportion of families with the sensitive character is 0.1.

In this study, we consider two differents sample sizes. First, 1000 samples are drawn by cluster sampling $m = 20$ and $m = 10$ cluster) by district. The same model parameters are considered in each model. Secondly, we consider a stratified cluster sampling. 1000 samples of 25 and 10 clusters are selected in the population using stratified sampling with proportional allocation.

To begin the study, we will observe the sample means of estimates of variances obtained by each randomization model (see Tables 1 and 2).

If we observe the different variances obtained by each randomization model, we observe that the estimates obtained by $\hat{V}(\hat{\hat{\overline{Y}}}_{\mathrm{HT}}(r))$ are close to the theoretical values $(V(\hat{\hat{\overline{Y}}}_{\mathrm{HT}}(r)))$. In the case of the variance obtained by the JK resampling technique,$\hat{V}_{\mathrm{JK}}(\hat{\hat{\overline{Y}}}_{\mathrm{HT}}(r))$, we see that it underestimates the true value of the variance; however, when we make the proposed adjustment, $\hat{V}^{ad}_{\mathrm{JK}}(\hat{\hat{\overline{Y}}}_{\mathrm{HT}}(r))$, it takes values very close to the empirical variances. These results are shown both in the simulation study and in the real population considering both samples sizes.

Tables 3 and 4 show the proportion of underestimation of variance, the absolute RB (|RB|), the RMSE and the coverage.

In Table 3, it is noted that the number of times that adjusted JK estimator underestimates the variance is considerably less than the usual estimator. There is also a decrease in the RB and the RMSE for all RR schemes. This decrease in bias and error is very noticeable in some models like the Warner model. This behavior is the same for the three populations considered. The coverage of intervals with the adjusted estimators is also larger.

Results for the families' population (see Table 4) demonstrated a similar pattern: the number of times that adjusted JK estimator underestimates the variance is considerably less than the usual estimator and the decrease in the realtive bias

TABLE 1 Variance for simulated populations

	Qualitative variables											
	$V\left(\hat{\hat{\overline{Y}}}_{\mathrm{HT}}(r)\right)$			$E\hat{V}\left(\hat{\hat{\overline{Y}}}_{\mathrm{HT}}(r)\right)$			$E\hat{V}_{\mathrm{JK}}\left(\hat{\hat{\overline{Y}}}_{\mathrm{HT}}(r)\right)$			$E\hat{V}^{ad}_{\mathrm{JK}}\left(\hat{\hat{\overline{Y}}}_{\mathrm{HT}}(r)\right)$		
	Pop I	Pop II	Pop III	Pop I	Pop II	Pop III	Pop I	Pop II	Pop III	Pop I	Pop II	Pop III
R_1	0.01211	0.01227	0.01238	0.01209	0.01225	0.01236	0.01149	0.01165	0.01176	0.01209	0.01225	0.01236
R_2	0.00095	0.00110	0.00121	0.00096	0.00112	0.00123	0.00092	0.00107	0.00119	0.00096	0.00112	0.00123
R_3	0.00098	0.00113	0.00125	0.00097	0.00112	0.00124	0.00093	0.00108	0.00119	0.00097	0.00112	0.00124

TABLE 2 Variance for the families population

	$V\left(\hat{\hat{Y}}_{HT}(r)\right)$	$E\hat{V}\left(\hat{\hat{Y}}_{HT}(r)\right)$	$E\hat{V}_{JK}\left(\hat{\hat{Y}}_{HT}(r)\right)$	$E\hat{V}_{JK}^{ad}\left(\hat{\hat{Y}}_{HT}(r)\right)$	$V\left(\hat{\hat{Y}}_{HT}(r)\right)$	$E\hat{V}\left(\hat{\hat{Y}}_{HT}(r)\right)$	$E\hat{V}_{JK}\left(\hat{\hat{Y}}_{HT}(r)\right)$	$E\hat{V}_{JK}^{ad}\left(\hat{\hat{Y}}_{HT}(r)\right)$
				Stratified sampling				
		n = 500					**n = 250**	
				Quantitative variables				
R_4	8849.1560	9502.5310	6176.2060	9502.9487	18092.4500	18799.0600	15473.8054	18796.2233
R_5	14280.1100	16084.0200	9766.3178	16084.4074	28812.5400	30926.5800	24589.9747	30922.2505
R_6	1509.2320	1655.2090	1275.2831	1655.4382	3389.1430	3561.0440	3178.3110	3557.6520
				Cluster sampling				
				Qualitative variables				
		m = 20					**m = 10**	
R_1	0.0203	0.0203	0.0160	0.0200	0.0408	0.0405	0.0361	0.0401
R_2	0.0011	0.0011	0.0009	0.0011	0.0033	0.0035	0.0030	0.0033
R_3	0.0018	0.0017	0.0015	0.0018	0.0036	0.0035	0.0033	0.0036
				Quantitative variables				
		m = 25					**m = 10**	
R_4	14280.1100	16084.0200	9766.3178	16084.4074	139399.4000	30787.2500	136802.9725	140100.2379
R_5	54935.8300	20943.5200	50523.1950	56828.5398	158161.1000	49983.2600	153397.4228	159698.4192
R_6	38709.2600	2289.2280	38087.7365	38467.4550	115453.0000	6107.0490	114248.5393	114628.1973
				Stratified cluster sampling				
		m = 25					**m = 10**	
				Qualitative variables				
R_1	0.0168	0.0167	0.0127	0.0167	0.0461	0.0456	0.0394	0.0434
R_2	0.0014	0.0014	0.0012	0.0015	0.0041	0.0040	0.0036	0.0039
R_3	0.0015	0.0014	0.0012	0.0015	0.0041	0.0040	0.0037	0.0039
				Quantitative variables				
R_4	50651.1300	14057.6700	48201.0747	51525.5914	165038.7000	46597.0800	151326.0482	154612.5416
R_5	57648.3300	22422.0900	52815.3255	59089.5568	186456.7000	69360.9200	170052.9597	176361.6367
R_6	40799.2700	3484.7980	40109.8589	40490.0023	138151.9000	20508.5200	127977.7802	128357.2015

TABLE 3 NB,|RB|, RMSE and COVE for simulated populations

		Qualitative variables					
		$\hat{V}_{JK}\left(\hat{\bar{Y}}_{HT}(r)\right)$			$\hat{V}_{JK}^{ad}\left(\hat{\bar{Y}}_{HT}(r)\right)$		
		Pop I	**Pop II**	**Pop III**	**Pop I**	**Pop II**	**Pop III**
R_1	NB	1.000	1.000	1.000	0.476	0.503	0.501
	|RB|	0.051	0.050	0.051	0.012	0.009	0.007
	RMSE	0.003	0.003	0.003	0.000	0.000	0.000
	COVE	0.946	0.950	0.942	0.949	0.957	0.955
R_2	NB	0.697	0.711	0.730	0.388	0.328	0.300
	|RB|	0.053	0.038	0.028	0.049	0.037	0.028
	RMSE	0.004	0.002	0.001	0.004	0.002	0.001
	COVE	0.934	0.939	0.935	0.943	0.939	0.946
R_3	NB	0.827	0.895	0.962	0.552	0.568	0.607
	|RB|	0.062	0.053	0.044	0.044	0.033	0.022
	RMSE	0.006	0.004	0.003	0.003	0.002	0.001
	COVE	0.939	0.940	0.950	0.945	0.940	0.963

and the RMSE for all RR schemes and for all sampling designs is noticeable. We obtain similar conclusions regardless of the sample size considered.

4.2 | Quantitative variables

For quantitative variables, we consider simulations from real populations. The first study considers the previous population of 1500 families to investigate their income tax return. The sample is drawn by stratified sampling by house ownership. We select 1000 stratified samples of size $n = 500$ and $n = 250$ with optimal allocation.

For this quantitative population, RR data were generated by using three different RR models R_t, $t = 4,5$ and 6. R_4 model is the Eichhorn and Hayre[28] model where the scramble variable is $S \sim F(20,20)$. R_5 model is the Bar-Lev, Bobovitch, and Boukai model proposed by Bar-Lev et al.[29] with the scramble variable $S \sim \exp(1)$, and probability $p = 0.6$. The third considered model R_6 is the Eriksson[30] model with parameters $Q = (3515, 4000, 4537), c = 0.7, q_1, q_2, q_3 = 0.1$.

In Table 2, if we consider the quantitative variables, we obtain the same conclusions that in the qualitative variables case. Through the adjusted JK variance, we can obtain good estimates of the empirical variance, considering both samples sizes, although we get better estimates when the sample size is larger.

For all models used for quantitative variable (see Table 4), the number of times that adjusted JK estimator underestimates the variance is considerably less than the usual estimator. There is also a decrease in the RB and the RMSE for all RR schemes, and this decrease in bias and error is as noticeable as in the case of qualitative variables when the sample size is larger. When we consider a smaller sample size, these values decrease but to a lesser extent.

For the quantitative variable, we made other simulation study using a more complex sampling. In this simulation, the samples are drawn by stratified cluster sampling. The sampling sizes and the parameters are the same as in the previous section.

The behavior of adjusted estimators against the original JK estimator is also the same: the number of times the proportion of nonpositive bias is low and the RB and the RMSE is also slightly low for the proposed estimator.

In summary, our simulation study shows that the proposed estimator is able to decrease the problem of the underestimation of the JK estimator of the variance, regardless of the sampling design and sample size considered. This behavior is valid for all RR schemes considered, although more evidences were found for the qualitative variables.

5 | A REAL EXAMPLE: ANALYZING BEHAVIORS AT THE UNIVERSITY

The usefulness of the proposed technique was illustrated in the context of students' behavior at the University of Granada. To investigate sensitive behaviors such as cheating in exams and bullying, we conducted a survey of students.

5.1 | Participants and sampling method

The target population for this survey was a group of students at the University of Granada. Subjects were selected using a complex sampling design in which a stratified (by university faculty) cluster sample was extracted such that degree subjects, and levels were represented in proportion to their total numbers of students. All questionnaires were administered during regular class-meeting times. According to local law, the study does not fall under the Human Research Act and hence does not require authorization from an ethics committee. Therefore, we do not have an approval from a formal Institutional Review Board for this study.

5.2 | Procedure and measure

The questionnaire began with a set of basic demographic questions, followed by some academic questions and then a number of items referring to sensitive behavior such as cheating in exams and bullying.

TABLE 4 NB,|RB|, RMSE and COVE for the families population

		$\hat{V}_{JK}\left(\hat{\overline{Y}}_{HT}(r)\right)$	$\hat{V}_{JK}^{ad}\left(\hat{\overline{Y}}_{HT}(r)\right)$	$\hat{V}_{JK}\left(\hat{\overline{Y}}_{HT}(r)\right)$	$\hat{V}_{JK}^{ad}\left(\hat{\overline{Y}}_{HT}(r)\right)$
		Stratified sampling			
		n = 500		**n = 250**	
		Quantitative variables			
R_4	NB	0.999	0.275	0.830	0.475
	|RB|	0.302	0.103	0.192	0.140
	RMSE	0.100	0.018	0.052	0.038
	COVE	0.895	0.955	0.933	0.955
R_5	NB	0.970	0.258	0.773	0.458
	|RB|	0.322	0.168	0.251	0.220
	RMSE	0.120	0.050	0.090	0.090
	COVE	0.871	0.959	0.924	0.952
R_6	NB	1.000	0.056	0.748	0.315
	|RB|	0.155	0.099	0.090	0.088
	RMSE	0.026	0.013	0.012	0.012
	COVE	0.919	0.939	0.937	0.952
		Cluster sampling			
		Qualitative variables			
		m = 20		**m = 10**	
R_1	NB	0.812	0.563	0.675	0.583
	|RB|	0.278	0.201	0.359	0.335
	RMSE	0.107	0.063	0.196	0.184
	COVE	0.904	0.927	0.894	0.918
R_2	NB	0.699	0.515	0.647	0.566
	|RB|	0.265	0.230	0.379	0.361
	RMSE	0.104	0.089	0.221	0.217
	COVE	0.911	0.938	0.906	0.923
R_3	NB	0.741	0.550	0.632	0.567
	|RB|	0.279	0.230	0.366	0.350
	RMSE	0.112	0.087	0.200	0.195
	COVE	0.912	0.934	0.900	0.915
		Quantitative variables			
		m = 25		**m = 10**	
R_4	NB	0.970	0.258	0.561	0.525
	|RB|	0.322	0.168	0.276	0.275
	RMSE	0.120	0.050	0.122	0.122
	COVE	0.871	0.959	0.915	0.918
R_5	NB	0.648	0.471	0.580	0.534
	|RB|	0.196	0.186	0.304	0.302
	RMSE	0.058	0.056	0.156	0.158
	COVE	0.921	0.940	0.905	0.908
R_6	NB	0.533	0.514	0.527	0.518
	|RB|	0.121	0.120	0.235	0.235
	RMSE	0.023	0.023	0.085	0.085
	COVE	0.923	0.927	0.916	0.916

(Continues)

TABLE 4 continued

		Stratified cluster sampling			
		m = 25		m = 10	
Qualitative variables					
R_1	NB	0.843	0.573	0.716	0.654
	|RB|	0.303	0.198	0.455	0.423
	RMSE	0.123	0.064	0.382	0.366
	COVE	0.909	0.940	0.900	0.926
R_2	NB	0.754	0.482	0.681	0.631
	|RB|	0.271	0.214	0.455	0.433
	RMSE	0.114	0.092	0.401	0.392
	COVE	0.910	0.946	0.908	0.925
R_3	NB	0.809	0.568	0.680	0.629
	|RB|	0.290	0.209	0.482	0.461
	RMSE	0.115	0.064	0.496	0.488
	COVE	0.913	0.946	0.862	0.882
Quantitative variables					
R_4	NB	0.630	0.511	0.648	0.635
	|RB|	0.193	0.184	0.401	0.396
	RMSE	0.061	0.059	0.395	0.392
	COVE	0.926	0.935	0.881	0.884
R_5	NB	0.680	0.508	0.669	0.645
	|RB|	0.220	0.205	0.418	0.409
	RMSE	0.074	0.071	0.357	0.354
	COVE	0.917	0.936	0.884	0.898
R_6	NB	0.554	0.526	0.647	0.646
	|RB|	0.151	0.150	0.397	0.396
	RMSE	0.039	0.039	0.338	0.338
	COVE	0.931	0.932	0.874	0.874

TABLE 5 Instructions to complete the questionnaire

HEADS	Have you ever cheated in an exam?	Yes	No
TAILS	Were you born in July?		
HEADS	Have you ever bullied someone?	Yes	No
TAILS	Does your ID number end with 5?		

To randomize the responses, we used the model proposed by Horvitz et al[23] and extended by Greenberg et al[24] (R_2 model). Compared to alternative variants of the RRT, this design is less complex and very suitable.

The randomizing device used was the app “Randomizers” with the “Coin Flipper” option, which was installed on the student's phone. The application is very easy to use: the user just touches the “Randomize” button and one side of the coin is shown, head or tail. In our study, if “head” appeared, the student answered the main question and if “tail” appeared, the alternative question was answered.

Table 5 shows the sensitive questions and their respective unrelated questions.

The teacher explained that this technique preserved the students anonymity; thus, the RR process did not provoke any mistrust: all the students completed the questionnaire.

Eight interviewers carried out the interviews from January to March, 2015, with a final sample of 710 students.

5.3 | Software

Standard software packages for complex surveys cannot be used directly when the sample is obtained using RR techniques. Although analyses with standard statistical software, with certain modifications in the randomized variables, can yield correct point estimates of population parameters, they could still yield incorrect results for the standard errors estimated. R packages have been developed for estimation with RR surveys, such as the RRreg package[31] and the rr package,[32] but the methods implemented in these packages assume simple random sampling. Therefore, we used the package RRTCS,[33] which is the only one that incorporates estimation procedures for handling RR data obtained from

TABLE 6 Sociodemographic distribution of the sample

	Frequency	Percentage
Total	710	100%
Gender		
Male	335	47.15%
Female	375	52.85%
Faculty		
Science. Engineering	225	31.7%
Health	154	21.7%
Social Sciences. Law	331	46.6%

TABLE 7 Estimated proportions and confidence intervals for the variables

Variables	Estimated proportion	Estimated variance	95% Confidence interval
Cheating	0.8432	0.0013944	(0.77, 0.91)
Bullied	0.1288	0.00056	(0.082, 0.175)

complex surveys. To construct the estimated variance and the confidence interval, we used the adjusted JK estimator for a stratified cluster sample.

5.4 | Results

The sociodemographic distribution of the sample is shown in Table 6.

Respondents were separated by gender and by university faculty. The sample was composed of 710 students, of whom 52.85% were women and 47.15% were men. By faculties, 31.7% of the students studied in the Science and Engineering faculty, 21.7% in the Health Sciences faculty and 46.6% in the Social Sciences and Law faculty.

The results were as follows (see Table 7).

The estimated prevalence of students who had ever bullied someone ranged from 8.2%to 17.5%(at the 95%confidence level).

The estimated prevalence of students who had cheated in an exam someone ranged from 77% to 91% (at the 95% confidence level). This survey shows that students cheat on exams become an important problem at our university. The cheating is remarkably higher than the previous RR surveys conducted by Scheers and Dayton[34] that underreporting of five academic cheating behaviors ranged from 39% to 83% when responses to an anonymous questionnaire were compared to estimates using an RRT; Fox[35] indicated that about 25% of the respondents have once cheated on an exam. However, almost 55% of the students admit that they have observed cheating; and Fox and Meijer,[36] who studied a set of items related to cheating, showed that items with high threshold values were endorsed only by students who cheat often, only 25% of all students, of which 72% were men and 49% women. The reason for this increase could be the introduction of mobile phones and laptop that has led to high-tech cheating and, on the other hand, that punishments for cheating in our university are very mild.

6 | DISCUSSION

The idea of perturbing the values of a sensitive variable by the use of scrambling distributions has been used in many randomization devices, with the goal of increasing respondents cooperation. There have been many reports that RR provides more accurate estimates of the prevalence of socially undesirable behavior than does asking the sensitive question directly.[37] Numerous empirical studies have shown that RR obtains higher estimates of sensitive characteristics than are produced by DQ.

A major drawback of RRT models is their low efficiency, that is, their greater sampling variance as compared to traditional surveys.[38] The present contribution addresses problems of biased variance estimation by using resampling methods when a RR survey is obtained from a complex survey design. We have shown that the JK variance estimator underestimate the variance as was reported by Wolter[14] for multistage sampling. We also proposed modification of the conventional JK variance estimator for complex survey designs including cluster and stratification. The performance of the proposed modified estimator is tested for six RR techniques and various sample sizes, with respect to the criteria: underestimation, RB, efficiency, and coverage through simulation studies. The simulation studies reveal that the number of times that adjusted JK estimator underestimates the variance is considerably less than the usual estimator. There is also a decrease in the RB and the RMSE for all RR schemes. This decrease in bias and mean square error is very noticeable for qualitative variables but for quantitative variables not to such extent. In this paper, we have considerered JK methods to variance estimation in RR but others resampling techniques as bootstrap can also be considered.

ACKNOWLEDGEMENT

This work is partially supported by Ministerio de Economía y Competitividad of Spain (Grant MTM2015-63609-R).

ORCID

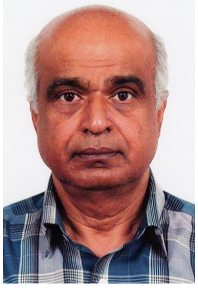
Beatriz Cobo https://orcid.org/0000-0003-2654-0032

REFERENCES

1. Chaudhuri A, Christofides TC. *Indirect Questioning in Sample Surveys*. Berlin, Germany: Springer-Verlag; 2013.
2. Warner SL. Randomized response: a survey technique for eliminating evasive answer bias. *J Am Stat Assoc*. 1965;60(309):63-69.
3. Perri PF, Rueda M, Cobo B. Multiple sensitive estimation and optimal sample size allocation in the item sum technique. *Biom J*. 2018;60:155-173.
4. Rueda M, Perri PF, Cobo B. Advances in estimation by the item sum technique using auxiliary information in complex surveys. *Adv Stat Anal*. 2018;102:455-478.
5. Tian G, Liu Y, Tang M. Logistic regression analysis of non-randomized response data collected by the parallel model in sensitive surveys. *Aust N Z J Stat*. 2019;61(2):134-151. https://doi.org/10.1111/anzs.12258
6. Chu AMY, So MKP, Chan TWC, Tiwari A. Estimating the dependence of mixed sensitive response types in randomized response technique. *Stat Methods Med Res*. 2019. https://doi.org/10.1177/0962280219847492
7. Rueda MM, Cobo B, López-Torrecillas F. Measuring inappropriate sexual behavior among university students: using the randomized response technique to enhance self-reporting. *Sex Abuse*. 2019. https://doi.org/10.1177/1079063219825872
8. Allende-Alonso S, Bouza-Herrera CN. Studying the quality of environment variables using a randomized response procedure for the estimation of a proportion through ranked set sampling. In: Bouza-Herrera CN, Al-Omari AIF, eds. *Ranked Set Sampling*. San Diego, CA:Academic Press; 2019:1-8.
9. Gibson MA, Gurmu E, Cobo B, Rueda MM, Scott IM. Indirect questioning methods reveal hidden support for female genital cutting in south Central Ethiopia. *PLoS ONE*. 2018;13(5):1-14.
10. Jing L, Lu Q, Cui Y, Yu H, Wang T. Combining the randomized response technique and the network scale-up method to estimate the female sex worker population size: an exploratory study. *Public Health*. 2018;160:81-86.
11. Kirtadze I, Otiashvili D, Tabatadze M, et al. Republic of Georgia estimates for prevalence of drug use: randomized response techniques suggest under-estimation. *Drug Alcohol Depend*. 2018;187:300-304.
12. Hsieh SH, Lee SM, Shen PS. Logistic regression analysis of randomized response data with missing covariates. *J Stat Plan Infer*. 2010;140(4):927-940.
13. Lensvelt-Mulders GJLM, Boeije HR. Evaluating compliance with a computer assisted randomized response technique: a qualitative study into the origins of lying and cheating. *Comput Hum Behav*. 2007;23(1):591-608.
14. Wolter KM. *Introduction to Variance Estimation*. New York, NY:Springer; 2007.
15. Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. *J Am Stat Assoc*. 1952;47:663-685.
16. Arnab R. Optional randomized response techniques for complex survey designs. *Biom J*. 2004;46(1):114-124.
17. Barabesi L, Diana G, Perri PF. Design-based distribution function estimation for stigmatized populations. *Metrika*. 2013;76:919-935.
18. Chaudhuri A. *Randomized Response and Indirect Questioning Techniques in Surveys*. Boca Raton, FL:Chapman & Hall; 2011.
19. Arcos A, Rueda M, Singh S. A generalized approach to randomised response for quantitative variables. *Qual Quant*. 2015;49:1239-1256.
20. Särndal CE, Swenson B, Wretman J. *Model Assisted Survey Sampling*. New York, NY:Springer; 1992.
21. Quenouille MH. Problems in plane sampling. *Ann Math Stat*. 1949;20:355-375.
22. Kim JM, Elam ME. A stratified unrelated question randomized response model. *Stat Pap*. 2007;48:215-233.
23. Horvitz DG, Shah BV, Simmons WR. The unrelated question RR model. *Proc ASA Soc Stat Sec*. 1967:65-72.
24. Greenberg BG, Abul-Ela AL, Simmons WR, Horvitz DG. The unrelated question RR model: theoretical framework. *J Am Stat Assoc*. 1969;64:520-539.
25. Boruch RF. Relations among statistical methods for assuring confidentiality of social research data. *Soc Sci Res*. 1972;1:403-414.
26. Fernández FR, Mayor JA. *Muestreo en Poblaciones Finitas: Curso Basico*. Barcelona, Spain: PPU; 1994.
27. Rueda M, González S, Arcos A. Indirect methods of imputation of missing data based on available units. *Appl Math Comput*. 2005;164(1):249-261.
28. Eichhorn B, Hayre LS. Scrambled randomized response methods for obtaining sensitive quantitative data. *J Stat Plan Infer*. 1983;7:307-316.
29. Bar-Lev SK, Bobovitch E, Boukai B. A note on randomized response models for quantitative data. *Metrika*. 2004;60:255-260.
30. Eriksson SA. A new model for randomized response. *Int Stat Rev*. 1973;41:40-43.
31. Heck DW, Moshagen M. Package RReg: correlation and regression analyses for randomized response data. 2015; http://psycho3.uni-mannheim.de/Home/Research/Software/RReg/

32. Blair G, Imai K, Zhou YY. Package rr: statistical methods for the randomized response. 2015; http://CRAN.R-project.org/package=rr
33. Cobo B, Rueda M. Arcos A. Package RRTCS: randomized response techniques for complex surveys. 2015; http://cran.r-project.org/web/packages/RRTCS/
34. Scheers NJ, Dayton CM. Improved estimation of academic cheating behavior using the randomized response technique. *Res High Educ*. 1987;26:61-69. https://doi.org/10.1007/BF00991933
35. Fox JO. Randomized item response theory models. *J Educ Behav Stat*. 2005;30(2):189-212. https://doi.org/10.3102/10769986030002189
36. Fox JP, Meijer RR. Using item response theory to obtain individual information from randomized response data: an application using cheating data. *Appl Psychol Measur*. 2008;32:595-610.
37. Fox JP, Wyrick C. A mixed effects randomized item response model. *J Educ Behav Stat*. 2008;33(4):389-415. https://doi.org/10.3102/1076998607306451
38. Ostapczuk M, Moshagen M, Zhao Z, Musch J. Assessing sensitive attributes using the randomized response technique: evidence for the importance of response symmetry. *J Educ Behav Stat*. 2009;34:267-287.

AUTHOR BIOGRAPHIES



Raghunath Arnab PhD, is a Professor of Statistics at the University of Botswana, Gaborone, Botswana, and an Honorary Professor of Statistics at the University of KwaZulu-Natal, Durban, South Africa. He received his PhD degree from the Indian Statistical Institute, Kolkata, India, in 1981. He is the author of the book "Survey Sampling Theory and Applications, Academic Press, UK." He is Associate editors of a few international journals and published more than 100 articles of international journals. He was an elected member of the International Statistical Institute and a Life member of the Indian Society of Agricultural Statistics.

Beatriz Cobo PhD, is currently a PhD Assistant Professor in Complutense University of Madrid, Madrid, Spain. She received her PhD degree in Mathematical and Applied Statistics from the University of Granada, Granada, Spain. Her professional interests include teaching and study of survey research methods, with particular emphasis on the indirect questioning techniques with the aim of obtaining more efficient estimators for sensitive questions, using, for example, auxiliary information and its computational treatment.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Arnab R, Cobo B. Variance jackknife estimation for randomized response surveys: A simulation study and an application to explore cheating in exams and bullying. *Comp and Math Methods*. 2020;2:e1073. https://doi.org/10.1002/cmm4.1073