PROCEDURES FOR SCRAMBLING SENSITIVE QUANTITATIVE VARIABLES: AN UPDATED REVIEW[1]

Pablo O. Juárez-Moreno[2]**, Carlos N. Bouza-Herrera*, Agustín Santiago-Moreno** and José M. Sautto-Vallejo**
*Facultad MATCOM, Universidad de La Habana, Cuba.
**Universidad Autónoma de Guerrero, México

ABSTRACT
In this document is presented a review of the current literature on randomized responses for quantitative variables. The review is chronological ordered considering its origin, Warner (1965), up to recent proposals. The techniques are classified and some are selected for classification. A study, using real data is developed using data that may be considered as sensitive, evaluating the accuracy and efficiency of the best quantitative randomized response techniques.

KEYWORDS: randomized responses, scrambling, simple random sampling, sensitivity level.

MSC: 62D05, 62P10

RESUMEN
Se presenta en este documento una revisión de la literatura sobre las técnicas de respuestas aleatorizadas (RR) para variables cuantitativas. Esta revisión se hace de forma cronológica tomando en cuenta su origen, Warner (1965), hasta técnicas RR cuantitativas recientes. Se realiza una clasificación de estas técnicas RR y se seleccionan algunas en cada clasificación. Se realiza un estudio con datos reales considerados sensibles para evaluar la precisión y eficiencia de las mejores técnicas RR cuantitativas.

PALABRAS CLAVE: respuestas aleatorizadas, codificar, muestreo aleatorio simple, nivel de sensibilidad.

1. INTRODUCTION

When sampling by survey, the main objective of the researcher is to know some characteristic of interest of a population. The aspiration of the sampler is that all the respondents in the sample answer the question of interest, but this hardly happens in reality, this lack of response is known as sampling errors. Such errors are caused, for example, by non-response or a false answer that is usually related to a sensitive question that respondents refuse to answer. To solve the above problem, Warner (1965) proposes a methodology which is mainly based on obtaining information about a sensitive characteristic such as drug-related issues, tax evasion, abortion, sexual crimes, alcoholism, etc., without giving know this information directly, that is, keeping the confidentiality of the respondent.
The essence of the Randomized Responses (RR) methods is as follows: an individual is interviewed to provide sensitive information Y, performs a randomized experiment whereby their response is scrambling. This response could be denoted by $R = f(Y, \theta)$, where the distribution of θ is known by the researcher. The response R is generated using a randomization device, say M. We can find its expectation $E_M(R)$ and variance $V_M(R)$ with respect to the device. We determine some functions whose inverses $g^{-1}\,(E_M(R)) = Y$ and $h^{-1}\,(V_M(R))$ are used to "un-scrambling". From a sample design d and since $\bar{y}(R) = \frac{1}{n}\sum_{i=1}^{n} R_i$ is the mean of the responses, its properties can be obtained through $E[\bar{y}(R)] = E_d E_M[\bar{y}(R)]$ and $V[\bar{y}(R)] = E_d V_M[\bar{y}(R)] + V_d E_M[\bar{y}(R)]$. Or failing that, by means of the central limit theorem.
Adapting the above notation to Warner's work (1965), the answer is given by $\pi_{yes} = f(\pi_A, P)$ with function $g^{-1}\,(\pi_{yes}) = \pi_A$, through which the value of the sensitive variable can be found, which in this case is π_A. The design defined by Warner (1965) is a SRS (d) so its mean and variance of the report π_{yes} can be known, $E[\hat{\pi}_A] = E_d E_M[\hat{\pi}_A]$ y $V[\hat{\pi}_A] = E_d V_M[\hat{\pi}_A] + V_d E_M[\hat{\pi}_A]$. We can call the latter the inference tools which in this document we focus on developing for each of the RR techniques presented.
The purpose of this work was to make a chronological and comparative review of the references of the quantitative RR methods that use SRSWR. To be more specific, the interest was to review only linear RR

---

1 **Presented at 15th International Conference on Operations Research, Havana (Cuba) May 31 – June 3, 2022**
2 18006@uagro.mx

methods in response models. In these methods, we will describe the scrambling device, specify the response model and estimator's characteristics of the population mean of the sensitive variable Y. On the other hand, in this paper we will not discuss generalized RR models, designs with RR, or qualitative RR methods, with the exception of Warner (1965). For a broader study of RR, it would be of interest to the reader to consult Chaudhuri et al. (2016).

In the first section, we present the pioneering work on RR of Warner (1965) and the early work based on his idea for quantitative data. In section two, a review of RR techniques for recent quantitative data is carried out, dividing them into: compulsory randomized response techniques (CRRT), full optional randomized response techniques (FORRT) and partial optional randomized response techniques (PORRT). In the third section, a comparative analysis is made between the best CRRT, FORRT and PORRT techniques. For comparison, a simulation study with sensitive type variables is performed. The first extended works of the Warner model (1965) were both, qualitative studies, Abdel-Latifet al. (1967), Horvitz et al. (1967), Greenberg et al. (1969), Moors (1971), as quantitative studies, Greengerg et al. (1971), Eriksson. (1973), Pollock and Beck (1976), Eichhorn and Hayre (1983).

1.1. The first work on RR: Warner 1965

As usual in RR documents, we begin with the study of the randomized response work carried out by Warner (1965). In which, a random sample s of size n is taken with simple random sampling with replacement (SRSWR) to estimate the population proportion π_A of a sensitive qualitative characteristic A. This feature is difficult to know through direct response, so Warner proposed to protect the privacy of the respondent through a scrambling device consisting of two packs of letters. The first cards pack with proportion $P(\neq 0.5)$ has the sentence "¿Do you belong to group A?", that is, the respondent is asked whether or not they have sensitive characteristic A, and the complement of A is the second pack of cards that have the sentence "¿Do you belong to group $\bar{A}$?". The procedure consists of the respondent selecting a card, once he sees his card (A ó $\bar{A}$), he/she (change in all the paper) informs the interviewer with the truthful answer "Yes" or "No", in turn, the interviewer is not aware of the type of card that the respondent selected so that the sensitive characteristic is not revealed to the interviewer and therefore protects their privacy. So that, the proportion of responses "Yes" is π_y , that is, the specified model is:

$$\pi_y = P\,\pi_A + (1-P)(1-\pi_A).$$

As the interest is to know the population proportion π_A of a sensitive characteristic A and knowing that ρ_{ys} is the total of Yes in the sample, we have:

i. $\hat{\pi}_A = \frac{\rho_{ys}-(1-P)}{2P-1}$, which is unbiased to π_A

ii. $Var(\hat{\pi}_A) = \frac{\pi_A(1-\pi_A)}{n} + \frac{P(1-P)}{n\,(2P-1)^2}$

iii. $\hat{V}ar(\hat{\pi}_A) = \frac{\rho_{ys}\,(1-\rho_{ys})}{(n-1)\ (2P-1)^2}$, which is an unbiased estimator of $Var(\hat{\pi}_A)$

1.2 First works in quantitative RR

In this section, we will review some of the early work for quantitative variables. Chronologically, the first contribution of RR methodology for quantitative data is given by Greenberg et al. (1971) since it means for quantitative variables, Ericksson (1973) presented the second quantitative RR work and them were followed by the work of Pollock et al. (1976) and lastly, we also analyze the work by Eichhorn and Hayre (1983). Next, we will describe the first quantitative techniques.

Greenberg, Kuebler, Abernathy and Horvitz technique (1971)

The work presented by Greenberg et al. (1971) was carried out ~~for~~ with the purpose of obtaining quantitative information regarding abortion and income. The RR device consisted of a sealed clear plastic box, which, for the abortion trial, has two questions printed on the box lid:

1. How many abortions have you had in your life?
2. If a woman has to work full time to earn a living, how many children does she think she should have?

Inside the box there are P_1 red balls with the number 1 and P_2 blue balls with the number 2. In which, the j-th respondent selects a ball with probability P_i and will answer either question $1(P_1)$ or question $2(1-P_1)$ without the interviewer realizing which ball was selected and thus, protecting their privacy of information. Regarding the topic of income, they used the same device only changing the questions according to the purpose of interest (abortion or income). We will focus only on abortion issue.

The interest parameter is μ_Y which is the population mean of the sensitive characteristic Y (question 1). For the estimation, two samples are taken using SRSWR from a population of size *N*. The response for the i-th respondent for the j-th sample is given by:

$$Z_{ji} = P_j\, Y_{ji} + (1 - P_j)X_{ji}$$

where Y_{ji} = the number of abortions the i-th respondent has had in the j-th sample with $E[Y_{ji}] = \mu_Y$ and $X_{ji=}$ the number of children a woman should be have if she works according to the i-th respondent in the j-the sample with $E[X_{ji}] = \mu_X$. Let us consider the expectation under the randomization procedure used to obtain answers from Z_{ji}:

$$E(Z_{ji}) = \mu_{zj} = P_j\, \mu_Y + (1 - P_j)\mu_X.$$

Now,

i. $\hat{\mu}_Y = \frac{(1-P_2)\bar{Z}_1 - (1-P_1)\bar{Z}_2}{P_1 - P_2}$, which is unbiased to μ_y and, $\bar{Z}_1$ and $\bar{Z}_2$ they are the sample means, $P_1 \neq P_2$

ii. $Var\,(\hat{\mu}_Y) = \frac{1}{(P_1-P_2)^2}\left[(1-P_2)^2\, Var(\bar{Z}_1) + (1-P_1)^2 Var\,(\bar{Z}_2)\right]$

Where, $Var(\bar{Z}_j) = \frac{1}{n_j}\left[\sigma_X^2 + p_j(\sigma_Y^2 - \sigma_X^2) + p_j(1-p_j)(\mu_Y - \mu_X)^2\right]$

iii. $\hat{V}ar\,(\hat{\mu}_Y) = \frac{1}{(P_1-P_2)^2}\left[(1-P_2)^2\, \hat{V}ar(\bar{Z}_1) + (1-P_1)^2 \hat{V}ar\,(\bar{Z}_2)\right]$, which is an unbiased estimator of $Var\,(\hat{\mu}_Y)$, Where, $\hat{V}ar(\bar{Z}_j) = \frac{1}{n_j(n_j-1)}\sum(Z_{ji} - \bar{Z})^2$, j=1,2.

Eriksson's technique (1973)

For this technique, a population of size N is considered whose sensitive characteristic is measured by $Y_1, Y_2, \dots, Y_N$, where the mean is denoted by μ_Y and its variance by σ_Y^2. The choice of L values $X_1, X_2, \dots, X_L$, similar, to the set of values $Y_1, Y_2, \dots, Y_N$, is assumed ~~in such a way as to~~ generates confidence in the interviewees that their information will not be known by the interviewer. The procedure to hide the data of the interviewees is described below.

A card deck is had which are of two types:

Type 1: ¡Provide your true answer!

Type 2: ¡Say your value is x_j!

In this card set there is a proportion P of the type 1 and in the case of cards of type 2, there is a proportion p_j for the card marked with the value X_j for $j = 1,2,..,L$ such that $\sum p_j = 1 - P$. For the type 2 cards, their means and variances are given by:

$$\mu_X = \sum X_j \frac{p_j}{1-P}, \quad \sigma_X^2 = \sum (X_j - \mu_X)^2 \frac{p_j}{1-P}.$$

With these considerations, the random variable Z_{iv} ~~is denoted~~, ~~which~~ represent the response of the i-th person in the *v*-th replicate, with probability $P, p_1, p_2, \dots, p_L$, respectively. There is a sample of interviewees of size *n*, and the different averages are defined as:

$$Z_{iv}\colon \bar{Z}_i = \frac{1}{k}\sum Z_{iv}\,, \bar{Z} = \frac{1}{kn}\sum\sum Z_{iv}.$$

It is assumed that a random sample of size *n* can be obtained directly and an estimator of μ_Y that is $\xi = \sum a_i Y_i$. In the case of randomized responses, a possible estimator of μ_Y is:

i. $\hat{\mu}_Y = \frac{1}{P}\sum a_i(\bar{Z}_i - (1-P)\mu_X)$

The variance of $\hat{\mu}_Y$ is given by:

ii. $V(\hat{\mu}_Y) = \frac{1}{kP^2}\sum a_i^2 V(Z_{iv})$

Where, $V(Z_{iv}) = (1-P)\left[\sigma_x^2 + P\left((Y_i - \mu_y)^2 + (\mu_y - \mu_x)^2 + 2(\mu_y - \mu_x)(Y_i - \mu_y)\right)\right]$

We propose as an estimator of $V(\hat{\mu}_y)$

iii. $\hat{V}(\hat{\mu}_y) = \frac{1}{kP^2}\sum a_i^2 \hat{V}(Z_{iv})$

Where, $\hat{V}(Z_{iv}) = (1-P)\left[\sigma_x^2 + P\left((Y_i - \hat{\mu}_y)^2 + (\hat{\mu}_y - \mu_x)^2 + 2(\hat{\mu}_y - \mu_x)(Y_i - \hat{\mu}_y)\right)\right]$

Pollock and Bek technique (1976)

Unlike the past techniques proposed by Greenberg et al. (1971) and Eriksson (1973), that use a binary model in questions ~~way~~ for scrambling the value of the sensitive characteristic Y, the technique presented by Pollock and Bek (1976) proposes another approach to scrambling that sensitive value Y by performing the addition or multiplication using a random value X from a known distribution. In the ~~first~~ addition model, the respondent is asked to sum his sensitive characteristic Y and a random value X, both values are ~~considered~~ independent and with distributions $f(y)$ and $g(x)$ respectively. So that, the response is given by:

$$Z = Y + X$$

With

$$E(Z) = \mu_z = \mu_Y + \mu_X \text{ and } Var(Z) = \sigma_z^2 = \sigma_Y^2 + \sigma_X^2.$$

We have the following characteristics of the estimator of μ_Y:

i. $\hat{\mu}_Y = \bar{Z} - \mu_X$, which is an unbiased estimator of μ_Y.
Further, $\bar{Z}$ is the estimator of μ_z

ii. $Var\ (\hat{\mu}_Y) = \frac{\sigma_Y^2+\sigma_X^2}{n}$

iii. $\hat{V}ar\ (\hat{\mu}_Y) \simeq \frac{2\,\sigma_X^2}{n}$

Pollock and Bek (1976) estimated the variance in this way, following Greenberg et al. (1971), whose recommend that X values should be similar to the Y values, so that the estimation is possible owing to the X values are known by the researcher.

The second model presented by Pollock and Bek (1976) is the multiplication, unlike the sum model, the value of sensitive characteristic Y of the respondent is scrambled by the multiplication of a X value. His response is given by:

$$Z = YX$$

with $E(Z) = \mu_z = \mu_Y\mu_X$ and $Var(Z) = \sigma_z^2 = \mu_X^2\sigma_Y^2 + \mu_Y^2\sigma_X^2 + \sigma_Y^2\sigma_X^2$.

The following characteristics of the estimator of μ_Y are presented:

i. $\hat{\mu}_Y = \frac{\bar{Z}}{\mu_X}$, is an unbiased estimator of μ_Y and $\bar{Z}$ is the estimator of μ_z

ii. $Var\ (\hat{\mu}_Y) = \frac{1}{n}\left[\sigma_Y^2 + \frac{\sigma_X^2(\mu_Y^2+\sigma_Y^2)}{\mu_X^2}\right]$

iii. $\hat{V}ar\ (\hat{\mu}_Y) \simeq \frac{1}{n}\left[2\,\sigma_X^2 + \frac{\sigma_X^4}{\mu_X^2}\right]$.

The variance estimator is specified following the same analogy of the sum model.

Eichhorn and Hayre technique (1983)

Eichhorn and Hayre (1983) works as follows. This technic~~,~~ as in the Pollock and Bek (1976) works as follows:

The respondent scrambles the sensitive value Y through the product of a positive random variable X with mean μ_X and variance σ_X^2 that are known by interviewer.

Therefore, the interviewer will obtain the report: $Z = YX$. Unlike the RR techniques already reviewed, where the sensitive value Y is scrambled through questions unrelated to the sensitive variable, the Eichhorn and Hayre technique (1983) is of particular interest since it can be considered the first technique where the respondent scrambles the sensitive value through a device, such as tossing a coin or extracting a ball from a bag, which are proposed by the authors.

The estimator of μ_Y present these characteristics:

i. $\hat{\mu}_Y = \frac{\bar{Z}}{\mu_X}$, is unbiased to μ_Y and $\bar{Z}$ is the estimator of μ_z

ii. $Var\ (\hat{\mu}_Y) = \frac{1}{n}\left[\sigma_Y^2 + \frac{\sigma_X^2(\mu_Y^2+\sigma_Y^2)}{\mu_X^2}\right]$, is the variance of the estimator

iii. $\hat{V}ar\ (\hat{\mu}_Y) = \frac{1}{n}\left[\frac{S_Z^2 - \frac{\sigma_X^2\hat{\mu}_Y^2}{\mu_X^2}}{\left(1+\frac{\sigma_X^2}{\mu_X^2}\right)}\right]$, which we propose as an estimator of $Var\ (\hat{\mu}_Y)$, with $S_Z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$,

After these first works on quantitative RR techniques, a large number of studies of these techniques are presented in the published literature produced before 21st century. We refer, among others, the papers of Chaudhuri (1987), Chaudhuri and Mukherjee (1988), Arnab (1995), Singh and Joarder (1997). In the 21st century, the RR techniques have been studied more thoroughly since their applicability are more frequent in different fields as: Medicine, Psychology, Biology, Social, Technology, and so on. The procedures have been called "scrambling". Some applied works from aforementioned areas, to mention a few, are those of Arnab (2004), Bouza (2009), Chaudhuri et al. (2009), Blair et al. (2015), Rueda et al. (2016). Also, the rise of RR techniques is due to their applicability in sensitive issues such as: abortion induction, see Lara et al. (2004) and Perri et al. (2016), drug use, see Stubbe et al. (2013) and Perry et al. (2017), AIDS see Arnab and Singh (2010), racism see Krumpal (2012), among others.

2. STUDY OF RECENT QUANTITATIVE RANDOMIZED RESPONSES TECHNIQUES

As we have said, the idea of using randomized responses techniques is to scramble the respondent's response through a device that implements RR in order to obtain information or the value of a sensitive characteristic of a person, who for some reason that person is refuses to answer. On the other hand, there are people who might consider that the question is non-sensitive or little sensitive, so they would be willing to respond directly to the question, that is, without the need to scramble their response. This is called optional randomized responses techniques (ORRT), see Chaudhuri and Mukherjee (1988), in which the respondent is selected by means of SRSWR and chooses one of the following two options:

1. Give the true response.
2. Scramble your true response through a device.

The fundamental characteristic of the ORRT techniques, see Gupta and Thornton (2002), Arnab (2004) for example, is that they provide minimum variance estimators with respect to the compulsory randomized responses techniques (CRRT) because the latter, being directly scrambled respondent's value, it increases the bias in the estimation. This is to say, as in the ORR techniques, the respondent has the option of giving his response to the sensitive characteristic directly, without being scrambled, then values collected by the researcher for that characteristic are more reliable, thus reducing he bias generated by using a RR device and a gain in estimation precision is obtained. For all of the above and adding that ORR techniques are currently a topic under study, see Chaudhuri et al. (2016), this document will address them. On the other hand, it is also of our interest to study ~~the~~ CRR techniques since their main advantage over the ORR techniques is that the respondents feel more confident since they do not enter the predicament of deciding whether to give their sensitive value or if scramble their value, that this is the opposite case to ORR techniques. We can categorize, see Arnab and Rueda (2016), the optional randomized responses techniques (ORRT) as: Full optional randomized response technique (FORRT) y Partial optional randomized response technique (PORRT). Therefore, we classify RR techniques into three categories, CRRT, FORRT y PORRT. Next, we describe and show recent RR works for each category.

2.1. Compulsory randomized response technique (CRRT)

Here, we will look at RR techniques where the respondent's response randomly is compulsorily scrambled by a device or forced to provide his true value. Unlike ORR techniques, here the respondent has no possibility to decide whether to provide his sensitive value or to scramble it. In these compulsories techniques (CRRT), when the respondent directly scrambled his response, he is more confident in protecting his privacy, this is, the interviewer will never doubt whether he is providing a direct response of his true value or a scrambled response. For the following techniques, the selection of the sample *s* of size *n* is through SRSWR.

2.1.1. Bar-Lev et al. (2004).

For the RR technique proposed by Bar-Lev et al. (2004), the i-th respondent performs a Bernoulli trial with probability P predefined by the interviewer. The value provided by the respondent will be given according to the Bernoulli trial α, by:

$$\alpha = \begin{cases} 1; \text{ the respondent gives his sensitive value Y} \\ 0; \text{ the respondent scrambles his value using } R_i = X_i Y_i \end{cases}$$

Where X_i is a random variable with mean μ_X and variance σ_X^2 which are known by the researcher. The response model for the i-th respondent is:

$$Z_i = \alpha Y_i + (1-\alpha) R_i$$

The estimator of μ_Z has the following characteristic:

i. $\hat{\mu}_Y = \frac{\bar{Z}}{P+Q\mu_X}$, which is an unbiased estimator of μ_Y and $\bar{Z}$ is the estimator of μ_z

ii. $Var\,(\hat{\mu}_Y) = \frac{1}{n}\left[\frac{\mu_Y^2\, C^2 + \sigma_Y^2\, C^2 - \mu_Y^2 U}{(P+Q\mu_X)^2}\right]$, where, $C^2 = Q\mu_X^2(1 + C_X^2) + P, \quad U = (P + \mu_X Q)^2, \quad C_X^2 = \frac{\sigma_X^2}{\mu_X^2}$

We develop the estimator of $Var\ (\hat{\mu}_Y)$,

iii. $\widehat{V}ar\ (\hat{\mu}_Y) = \frac{1}{n}\left[\frac{\hat{\mu}_Y^2\, C^2 + S_Y^2\, C^2 - \hat{\mu}_Y^2 U}{(P+Q\mu_x)^2}\right]$, where $S_Y^2 = \frac{1}{n}\left[\frac{(P+Q\mu_x)^2\, S_Z^2 - \hat{\mu}_Y^2\, C^2 + \hat{\mu}_Y^2 U}{C^2}\right]$,
with $S_Z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$

2.1.2. Ryu et al. (2006)

In the Ryu et al.'s model (2006), the authors consider a two-stage model. In the same way as the previous technique, in the first stage a Bernoulli trial α with probability P_1 is performed which the i-th respondent provides his sensitive value Y or goes to the next stage according to α:

$$\alpha = \begin{cases} 1; \text{ the respondent gives his sensitive value } Y \\ 0; \text{ the respondet goes to the next stage} \end{cases}$$

Again, for this second stage, a Bernoulli trial β with probability P_2 is performed, in such a way that:

$$\beta = \begin{cases} 1; \text{ the respondent gives his sensite value } Y \\ 0; \text{ the respondent scrambles with X Y} \end{cases}$$

Where X is a random variable with distribution known by the researcher and assuming $\mu_X = 1$. The response model is given by:

$$Z_i = \alpha Y_i + (1-\alpha)\beta Y_i + (1-\alpha)(1-\beta)X_i Y_i$$

The unbiased estimator of μ_Y is:

i. $\hat{\mu}_Y = \frac{\sum_{i=1}^{n} Z_i}{n}$,

ii. $Var\ (\hat{\mu}_Y) = \frac{V(Z_i)}{n} = \frac{\sigma_Y^2 + Q_1 Q_2 \sigma_X^2\ (\mu_Y^2 + \sigma_Y^2)}{n}$, which is the variance of the estimator

For $Q_t = 1 - P_t, t = 1,2$ we propose as an estimator of $Var\ (\hat{\mu}_Y)$,

iii. $\widehat{V}ar\ (\hat{\mu}_Y) = \frac{S_Y^2 + Q_1 Q_2 \hat{\sigma}_X^2\ (\hat{\mu}_Y^2 + S_Y^2)}{n}$, where $S_Y^2 = \frac{S_Z^2 - Q_1 Q_2 \hat{\sigma}_X^2\ \hat{\mu}_Y^2}{(1 + Q_1 Q_2 \hat{\sigma}_X^2)}$, with $S_Z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$

2.1.3. Saleem I. et al. (2019)

The authors, Saleem I. et al. (2019), in this technique they propose an RR model to directly scramble the sensitive variable Y through addition, a technique proposed by Pollock and Beck (1976), subtraction, proposed by Hussain (2012) or multiplication, proposed by Eichhorn and Hayre (1983). That is, the respondent's response is scrambled using one of three aforementioned models. The report is given by:

$$Z = g(Y + \alpha S) + (1-g)Y\,S,$$

Where S is the scrambling random variable with mean $\bar{S} = E(S) = 0$ and variance σ_S^2 and with constants known by the researcher $g \in [0,1]$ and $\alpha \in [-1,1]$. So the response is given by one of these scrambling:

$$Z = \begin{cases} Y + S & if\ \ g = 1\ and\ \alpha = 1 \\ Y - S & if\ \ g = 1\ and\ \alpha = -1 \\ Y\,S & if\ \ g = 0 \end{cases}$$

With μ_Y presenting the following characteristics:

i. $\hat{\mu}_Y = \bar{Z}$, which is unbiased to μ_Y

ii. $V(\hat{\mu}_Y) = \frac{1}{n}[g^2(\sigma_Y^2 + \alpha^2\sigma_S^2) + (1-g)^2\sigma_S^2(\bar{Y}^2 + \sigma_Y^2) + 2\alpha g(1-g)\bar{Y}\sigma_S^2]$

The proposed variance estimator of $(\hat{\mu}_Y)$ is:

iii. $\widehat{V}ar\ (\hat{\mu}_Y) = \frac{1}{n}\left[g^2\left(\hat{S}_Y^2 + \alpha^2\sigma_S^2\right) + (1-g)^2\sigma_S^2\left(\hat{\mu}_Y{}^2 + \hat{S}_Y^2\right) + 2\alpha g(1-g)\hat{\mu}_Y\sigma_S^2\right]$,
where $\hat{S}_Y^2 = \frac{\hat{S}_Z^2 - g^2\alpha^2\sigma_S^2 - (1-g)^2\sigma_S^2\hat{\mu}_Y{}^2 - 2\alpha g(1-g)\hat{\mu}_Y\sigma_S^2}{(g^2 + (1-g)^2\sigma_S^2)}$, with $\hat{S}_Z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$

2.1.4. Bouza et al. (2022)

In the procedure of the RR technique proposed by Bouza et al. (2022), the i-th sample respondent ~~s~~ performs the following two-stage procedure. In stage one, the respondent performs a Bernoulli trial γ with probability P. For stage two, the respondent provides his true value scrambled according to γ, so the report is:

$$Z_i = \begin{cases} S_i & if \quad \gamma_i = 1 \\ T_i & if \quad \gamma_i = 0 \end{cases}$$

Where the sensitive variable Y will be scrambled either by one of the following two devices for the i-th respondent:

$$S_i = Y_i + A_i \qquad \text{or} \qquad T_i = Y_i + B_i A_i \;;$$

where A_i and B_i are random variables with mean μ_A ,μ_B, and variance σ_A^2 ,σ_B^2, respectively, which are controlled by the researcher.

The report of the i-th respondent is modeled by:

$$Z_i = \gamma_i S_i + (1-\gamma_i) T_i.$$

The estimator of μ_Y has the following characteristics:

i. $\hat{\mu}_Y = \bar{Z} - \mu_A(P + Q\mu_B)$, which is an unbiased estimator of μ_Y and,

ii. $V(\hat{\mu}_Y) = \frac{\sigma_Y^2 + \sigma_A^2(P+Q\sigma_B^2)}{n}$

We present as a proposal for the variance estimator $(\hat{\mu}_Y)$:

iii. $\hat{V}ar\,(\hat{\mu}_Y) = \frac{\hat{\sigma}_Y^2 + \sigma_A^2(P+Q\sigma_B^2)}{n}$,

where $\hat{\sigma}_Y^2 = \frac{S_Z^2 - \sigma_A^2(P+Q\sigma_B^2)}{n}$, with $S_Z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$

2.2. Full optional randomized response technique (FORRT) to quantitative data

The FORR techniques were proposed by Chaudhuri and Mukherjee (1988), in which respondents from the sample s of size n selected with SRSWR are optionally grouped into two groups. One group will be made up of those respondents who choose to respond directly to the sensitive question, as they considered it to be little or not at all sensitive. In the other group, there will be the respondents who choose to respond through a scrambling device, since they evaluated the question as sensitive, so the researcher assumes that in this group G they present the sensitive characteristic Y. Otherwise, for the other group $\bar{G}$ which are assumed to have the characteristic $\bar{Y}$. Below are some of the procedures of FORRT.

2.2.1. Huang (2008)

In the technique propose by Huang (2008), two samples s_1 and s_2 extracted by the SRSWR design are considered independent~~ly~~ and with size n_1 and n_2, respectively, where $n = n_1 + n_2$. Being an optional RR technique, the i-th element in samples s_1 and s_2 can choose between:

i) Give the sensitive value, or

ii) Scramble the sensitive value with $Z_j = X_j Y$.

Where X_j is a random variable with distribution known by the researcher with mean $\mu_{Xj} = 1$ and variance γ_{Xj} for the samples $s_j, j = 1{,}2$. Huang (2008) defines W_G as the proportion of individuals belonging to the sensitive group G, that is, who choose to scramble their responses. The interviewer fixes that proportion with mean μ_G and variance σ_{yG}^2. The i-th respondent in the j-th sample reports their response scrambled as Z_{ji} and let $\bar{z}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} z_{ji}$.

We present the characteristics of the estimator of μ_Y as we have been doing.

i. $\hat{\mu}_Y = \beta\,\bar{z}_1 + (1-\beta)\,\bar{z}_2$, which is unbiased to μ_Y and with $0 < \beta < 1$.

ii. $V(\hat{\mu}_Y) = \beta^2\,\frac{\sigma_{z1}^2}{n_1} + (1-\beta)\frac{\sigma_{z2}^2}{n_2}$

and the error is estimated by

iii. $\hat{V}(\hat{\mu}_Y) = \beta^2\,\frac{s_{z1}^2}{n_1} + (1-\beta)\frac{s_{z2}^2}{n_2}$, where $s_{zj}^2 = \frac{1}{(n_j)}\sum_{i\in s_j}\left(z_{ij} - \bar{z}_j\right)^2, j = 1{,}2$

2.2.2. Arnab (2018) – a

In Arnab's work (2018) are presented two PORR techniques, where the first one, Arnab (2018) – a will be developed here, and in the following section we will present the second technique, Arnab (2018) – b. The author converts the PORR techniques presented by Gupta et al. (2002) in FORR techniques. Like the other full optional RR techniques, a sample s is selected through SRSWR where the i-th element selected from the sample has the option of giving a direct response y_i, that is, if belonging to the non-sensitive group $\bar{G}$, is true or give a response scrambled by $Y_i \frac{X_i}{\mu_X}$ where X_i is a random sample from a distribution with known parameters μ_X and σ_X^2. The i-th report is given by:

$$Z_i = \begin{cases} Y_i & \text{for i} \in \bar{G} \\ Y_i\, S_i & for \text{ i} \in \text{G} \end{cases};$$

for $S_i = \frac{X_i}{\mu_X}$. The response model is:

$$Z_i = \beta_i\, Y_i\, S_i + (1 - \beta_i)\, Y_i$$

Where β_i=0 if i $\in \bar{G}$ or β_i=1 if i $\in G$.
The estimator of μ_Y and its characteristics are:

i. $\hat{\mu}_Y = \bar{z} = \frac{\sum_{i=1}^{n} z_i}{n}$, is unbiased for μ_Y

ii. $V(\hat{\mu}_Y) = V(\bar{z}) = \frac{\bar{\sigma}_Z^2}{n}$, as the estimator variance. Where $\bar{\sigma}_z^2 = \sigma_y^2 + C_x^2 W_G \mu_{yG}^2 \left(1 + C_{yG}^2\right)$.

iii. $\hat{V}(\hat{\mu}_Y) = \frac{\sum_{i=1}^{n}(z_i - \bar{z})^2}{n\,(n-1)}$, which is an unbiased estimator of the $V(\bar{z})$

2.2.3. Arnab (2018)-b

Like the previous technique, ~~here~~ Arnab (2018) converts a PORR technique proposed by Huang (2010) to FORRT. Through SRSWR the i-th respondent from the sample s_j is selected, for j=1,2, and is asked to provide the sensitive value Y_i if it is considered that the question is not sensitive, otherwise, the respondent must scramble the response Y_i with:

$$S_{ij} = \frac{X_j}{\mu_{Xj}}\, Y_i + T_j$$

where X_j and T_j are random variables with known means μ_{Xj}, μ_{Tj} and variances σ_{Xj}, σ_{Tj} defined by the researcher. The i-th report in the j-th sample is given by:

$$Z_{ij} = \alpha_i S_{ij} + (1 - \alpha_i) Y_i$$

where

$$\alpha_i = \begin{cases} 0 & for \text{ i} \in \bar{G} \\ 1 & for \text{ i} \in \text{G} \end{cases}.$$

The estimator properties of μ_Y are:

i. $\hat{\mu}_Y = \frac{\mu_{T2}\ \bar{Z}_1 - \mu_{T1}\ \bar{Z}_1}{\mu_{T2} - \mu_{T1}}$, is unbiased to μ_Y

ii. $V(\hat{\mu}_Y) = \frac{1}{(\mu_{T2} - \mu_{T1})^2}\left(\mu_{T2}^2 \frac{\sigma_{S1}^2}{n_1} + \mu_{T1}^2 \frac{\sigma_{S2}^2}{n_2}\right)$, is the estimator variance $\hat{\mu}_Y$

iii. $\hat{V}(\hat{\mu}_Y) = \frac{1}{(\mu_{T2} - \mu_{T1})^2}\left(\mu_{T2}^2 \frac{\hat{s}_{S1}^2}{n_1} + \mu_{T1}^2 \frac{\hat{s}_{S2}^2}{n_2}\right)$, which is an unbiased estimator of $V(\hat{\mu}_Y)$ and $\hat{s}_{sj}^2 = \frac{1}{(n_j)} \sum_{i \in s_j} \left(s_{ij} - \bar{s}_j\right)^2$

2.3. Partial optional randomized response technique (PORRT) to quantitative data

The main ideas of PORR techniques were given by Mangat and Singh (1994). In these techniques, the respondent has the option to answer the sensitive question directly or with an RR device depending on the sensitivity level W of the question. This means that, W is assumed to be the probability that all the respondents have of supplying the values of the sensitive characteristic using RR.

2.3.1. Gupta et al. (2002)

One of the first works of PORRT is the one presented by Gupta et al. (2002), where the respondent is selected using the SRSWR design. The interviewer gives the i-th respondent, from a sample s, the choice of providing their true response Y with probability 1-W or providing a scrambled ~~response with~~ SY ~~and probability W~~, where S is a random variable with mean μ_S and variance σ_S^2 both known. Y is the sensitive characteristic of the respondent ~~to know~~ and W is the sensitivity level of the question. The response model for the i-th respondent is given by:

$$Z = S^{\alpha} Y$$

for $\alpha = 1$ if the response is scrambled or $\alpha = 0$, if the respondent provides his true value.
The estimator characteristic of the parameter μ_Y are the following:

i. $\hat{\mu}_Y = \frac{\sum_{i=1}^{n} z_i}{n} = \bar{z}$, which is unbiased to a μ_Y

ii. $V(\hat{\mu}_Y) = \frac{1}{n}\left[\sigma_y^2 + W\sigma_S^2(\sigma_Y^2 + \mu_y^2)\right]$, as the estimator variance

iii. $\hat{V}(\hat{\mu}_Y) = \frac{1}{n}\left[S_y^2 + \widehat{W}\sigma_S^2(S_y^2 + \hat{\mu}_y^2)\right]$, which is an unbiased estimator to $V(\hat{\mu}_Y)$, where $S_Y^2 = \frac{S_Z^2 - \widehat{W}\sigma_S^2\hat{\mu}_y^2}{1+\widehat{W}\sigma_S^2}$ with $S_Z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$ and $\widehat{W} = \frac{\frac{1}{n}\sum_{i=1}^{n}\log(Z_i) - \log(\frac{1}{n}\sum_{i=1}^{n} Z_i)}{\delta}$, where $\delta = E[\log(S)]$

2.3.2. Singh and Gorey (2018)

For this partial optional RR technique, it is assumed, like all techniques already presented, the extraction from a sample *s* of size *n* using SRSWR. The authors proposed a two-stage model to estimate the population mean μ_Y of the sensitive random variable Y>0. We have the scrambling random variable S>0 with distribution known by the researcher with parameters μ_S and σ_S^2. Both random variables are considered mutually independent. The two-stage model is as follow:

Stage 1: In this stage, the elements of the sample *n* are divided into two groups, T is the proportion of the first group in which the respondent will give the sensitive value Y directly and reliably.

Stage 2: Here, the other group with proportion (1-T) is considered, where if the i-th respondent considers that the question is sensitive, and responds with a scrambled device Y $\frac{s}{\mu_S}$, otherwise, the i-th respondent directly reports the sensitive value Y.

In the Z response model, providing the direct value Y has probability T+(1-T) (1-W) and providing the scrambled value Y $\frac{s}{\mu_S}$ has probability (1-T) * W. Where W is the probability of using RR or also known as the sensitivity level of the question and which is assumed to be known. As in all the RR procedure, the respondent's privacy is protected as the interviewer doesn't know from which stage the respondent's response is provided. The response model is given by:

$$Z = (Y^V)\left\{Y\left(\frac{S}{\mu_S}\right)^U\right\}^{1-V}$$

for,

$$V = \begin{cases} 1; & \text{if the respondent gives a direct response in the first stage} \\ 0; & \text{if the respondent goes to the second stage} \end{cases}$$

In the second stage,

$$U = \begin{cases} 1; & \text{if the respondent scrambles the response} \\ 0; & \text{if the respondent gives the direct response} \end{cases}$$

U and V are independent with Bernoulli distribution and parameter W and T respectively. Following, we show the characteristics of the μ_Y estimator and the W estimator:

i. $\hat{\mu}_Y = \frac{1}{n}\sum_{i=1}^{n} Z_i = \bar{Z}$ is an unbiased estimator of μ_Y and $\bar{Z}$ is the estimator of μ_z

ii. $Var\,(\hat{\mu}_Y) = \frac{1}{n}\left[\sigma_Y^2 + \frac{\sigma_S^2}{{\mu_S}^2}(\sigma_Y^2 + \mu_Y^2)(1-T)W\right]$, which is estimator variance

iii. $\widehat{V}ar\,(\hat{\mu}_Y) = \frac{1}{n}\left[S_Y^2 + \frac{\sigma_S^2}{{\mu_S}^2}(S_Y^2 + \hat{\mu}_Y^2)(1-\hat{T})\widehat{W}\right]$.

Where, $S_Y^2 = \frac{\left(S_Z^2 - \widehat{w}(1-\hat{T})\frac{\sigma_S^2}{{\mu_S}^2}\hat{\mu}_Y^2\right)}{\left(1+\widehat{w}\frac{\sigma_S^2}{{\mu_S}^2}(1-\hat{T})\right)}$, with $S_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Z_i - \bar{Z})^2$, is an unbiased estimator of σ_Z^2.

iv. $\widehat{W} = \frac{\frac{1}{n}\sum_{i=1}^{n}\log(Z_i) - \log(\frac{1}{n}\sum_{i=1}^{n} Z_i)}{(1-T)(\delta - \log\theta)}, \delta \neq \theta \text{ and } T \neq 1$; where $\delta = E[log(S)]$

2.3.3. Hussain, Z. and Shahid, M., I. (2019)

We have cataloged this RR technique as a PORR technique, although this is not entirely true, since it contemplates a first scrambling of the respondents' responses directly having the option to decide, so the respondents "decide" how to give their response (direct o scrambled). Hussain, Z. and Shahid, M., I. (2019), proposed a two-stage RR model in which two independent samples s_1 and s_2 are selected through SRSWR. The procedure is the following:

Stage 1: A proportion F of respondents in the j-th sample are asked to provide the addition and subtraction of their sensitive value Y with two scrambled variables, this is, $Y + X_j - V_j$, where X, V are random variables with mean and variance known by the researcher.

Stage 2: Here, the remaining proportion of respondents $(1-F)$ are provided with a PORR technique, in which they can decide whether to report their true value Y or to report their sensitive value scrambled in the same way as in stage 1.

The probability of giving the sensitive response directly is $(1-F)(1-W)$ and the probability that the i-th respondent in the j-th sample scrambles the value is $F+(1-F)W$. The report response in the j-th sampled person is given by:

$$Z_j=\beta(Y+X_j-V_j)+(1-\beta)\{\alpha(Y+X_j-V_j)+(1-\alpha)X\},\quad j=1,2$$

Where α and β are Bernoulli variables with means W and F respectively.

The characteristics of the estimator of μ_Y and the estimator of W are the following:

i. $\hat{\mu}_Y=\frac{\bar{Z}_2(\mu_1-1)-\bar{Z}_1(\mu_2-1)}{\mu_1-\mu_2}$, is the unbiased estimator of μ_Y , $\mu_1\neq\mu_2$

ii. $Var\ (\hat{\mu}_Y)=\frac{1}{(\mu_1-\mu_2)^2}\left[(\mu_2-1)^2\left(\frac{\sigma_{Z_2}^2}{n_1}\right)+(\mu_1-1)^2\left(\frac{\sigma_{Z_2}^2}{n_2}\right)\right]$, variance of $\hat{\mu}_Y$ where $\mu_1\neq\mu_2$

The proposed estimator of the above ~~property~~ is:

iii. $\widehat{V}ar\ (\hat{\mu}_Y)=\frac{1}{(\hat{\mu}_1-\hat{\mu}_2)^2}\left[(\hat{\mu}_2-1)^2\left(\frac{\hat{\sigma}_{Z_2}^2}{n_1}\right)+(\hat{\mu}_1-1)^2\left(\frac{\hat{\sigma}_{Z_2}^2}{n_2}\right)\right]$, where $\hat{\sigma}_{Z_2}^2=\frac{1}{n_2-1}\sum_{i=1}^{n_2}\left(Z_i-\overline{Z}\right)^2$

$iv.\widehat{w}=\frac{1}{(1-F)}\left[\frac{\bar{Z}_2-\bar{Z}_1}{\mu_1-\mu_2}-F\right]$, is the unbiased estimator of W, $\mu_1\neq\mu_2,\ F\neq1$

$v.Var\ (\widehat{w})=\frac{1}{(1-F)^2(\mu_1-\mu_2)^2}\left[\left(\frac{\sigma_{Z_2}^2}{n_1}\right)+\left(\frac{\sigma_{Z_2}^2}{n_2}\right)\right]$, variance of $\widehat{w}$, $\mu_1\neq\mu_2,\ F\neq1$

where,

$\sigma_{Z_j}^2=\sigma_Y^2+(F+(1-F)W)(\delta_j^2+\gamma_j^2)+(\mu_j-1)^2(F+(1-F)W)[1-(F+(1-F)W],\ j=1,2$

3. A SIMULATION STUDY. COMPARASON BETWEEN CRRT, FORRT AND PORRT TECHINIQUES

Form the techniques presented in the previous section, we selected the most recent RR techniques from each classification to compare their efficiency and accuracy. Therefore, the RR techniques to be compared in this section are RRT_1 = Bouza et al. (2022) for CRRT, RRT_2 = Arnab (2018)-b for FORRT and RRT_3 = Singh and Gorey (2018) for PORRT, although the latter is not among the most recent, but it is the one that best meets the definition of a partial optional RR technique.

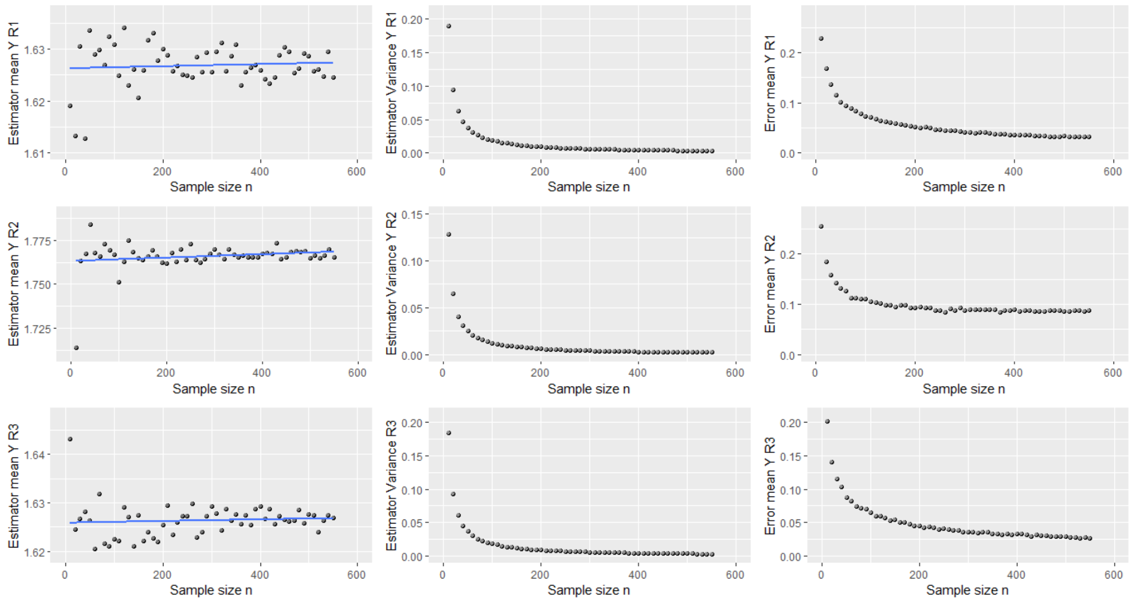
The data used for the simulation are those obtained from the National Survey on the Dynamics of Relationships in Households (2016), which among other aspects, presents information on the violence experienced, both in frequency and magnitude, by~~e~~ Mexican women over 15 years old in different areas and according to the relationship with their aggressors(s). Due to its social relevance and because we consider it to be of a sensitive nature, our population is the N=546 women who responded to the question "In your job, have they tried to force you to have sex against your will?" with an ordinal scale of: It did not occur = 1, Once = 2, Few times = 3 and Many times = 4, with parameters $=1.626$ and $\sigma^2=1.055$.

To evaluate the precision and efficiency for each sample, the following statistics were calculated:

$$Error(RRT)_s=\left(\frac{\frac{|\hat{y}_i-\bar{Y}_i|}{\bar{Y}_i}}{\frac{|\hat{y}_j-\bar{Y}_j|}{\bar{Y}_j}}\right)_s,\quad E(RRT)_s=\left(\frac{V(\bar{y}_i)}{V(\bar{y}_j)}\right)_s,$$

for the techniques, to evaluate i, j=1,2,3 and $i\neq j$. Given the population sampling error, the sample size was calculated, resulting in n=349 for SRSWR. A simulation of 1000 iterations was carried out and the aforementioned statistics were averaged. For the respective procedure, a probability P of belonging to the sensitive group of 0.7 and of not belonging to the sensitive group of 1-P=0.3 was assigned, since in a study of these characteristics the proportion of the sensitive group is always expected to be higher. The values of the scrambling variables used in the RR techniques to be compared, following Greenberg et al. (1971), are similar to the values of the sensitive characteristics Y.

Figure 1. Increase in sample size in RR techniques.



The results of the three procedures are presented below, both their estimator precision of the population mean and the variance efficiency of the mean estimator. See Tables 1 and 2.

From the results in Table 1, it can be seen that RRT_1 is more accurate than RRT_2 but RRT_3 is better than RRT_1 and RRT_2. In terms of efficiency, we can see in Table 2, the estimator with the lowest variance is RRT2, followed by RRT_3 and RRT_1. To visualize the behavior of the estimators of the three RR techniques, the following Figure 1 shows the graphs of $\hat{\mu}_{y_s}$, $\hat{V}ar\,(\hat{\mu}_Y)_s$ and $Error(RRT)_s$,where the samples sizes in the simulation were increasing from $n = 10{,}20, \dots , 550$.

Table 1. Accuracy of the estimators of the mean in the designs.

$\frac{Error(RRT_1)}{Error(RRT_2)} =$	0.453111956
$\frac{Error(RRT_1)}{Error(RRT_3)} =$	1.172517252
$\frac{Error(RRT_2)}{Error(RRT_3)} =$	2.58769877

~~As in~~ the numerical results of the simulation, in the graph is observed illustrates how ~~that~~ in general terms the RRT_3 technique is the best in precision over RRT_1 and RRT_2 since its estimates, with respect to the true parameter $\mu = 1.626$ are closer. This agrees with the graphs of the $Error(RRT)_s$. On the other hand, although numerically RRT_2 has a smaller variance than the other techniques, the researcher using these techniques would sacrifice precision with respect to RRT_3 and RRT_1.

Table 2. Efficiency of the variances of the estimators of the mean in the designs.

$\frac{E(RRT_1)}{E(RRT_2)} =$	1.439723845
$\frac{E(RRT_1)}{E(RRT_3)} =$	1.034732824
$\frac{E(RRT_2)}{E(RRT_3)} =$	0.71870229

ACKNOWLEDGMENTS: The authors give their heartily thanks to the referees´ invaluable comments, which allowed to improve the quality of the early versions of this paper. C. N. Bouza acknowledges the support of CITMA project PN223LH010-005.

RECEIVED: MAY, 2022.
REVISED: APRIL, 2023.

REFERENCES

[1] ABDEL-LATIF A. ABUL-ELA, BERNARD G. GREENBERG and DANIEL G. HORVITZ (1967)~~,~~ A Multi-Proportions Randomized Response Model. **Journal of the American Statistical Association**, 62, 319, 990-1008

[2] ARNAB, R., (1995): Optimal estimation of a finite population total under randomized response surveys. **Statistics** 27, 175–180.
[3] ARNAB, R., (2004). Optional randomized response techniques for complex survey designs. **Biom. J**. 46, 114–124.
[4] ARNAB, R. (2018). Optional randomized response techniques for quantitative characteristics, **Communications in Statistics - Theory and Methods**, DOI: 10.1080/03610926.2018.1489554.
[5] ARNAB, R., and M. RUEDA. (2016). Optional randomized response: A critical review. In Vol. 34 of **Handbook of Statistics**, ed. A. Chaudhuri, T. C. Christofides and C. R. Rao, 253–71. Oxford, UK: Elsevier.
[6] ARNAB, R. and SINGH, S. (2010). Randomized response techniques: An application to the Botswana AIDS impact survey, **J. Statist. Plann. Inference** 140, 941–953.
[7] BAR-LEV, S.K., BOBOVITCH, E., BOUKAI, B., (2004): A note on randomized response models for quantitative data. **Metrik**a 60, 255–260.
[8] BLAIR, G., IMAI, K., ZHOU, Y.Y., (2015): Package 'rr': Statistical Methods for the Randomized Response**. http://CRAN.R-project.org/package= rr**.
[9] BOUZA, C., N. (2009): Ranked set sampling and randomized response procedures for estimating the mean of a sensitive quantitative character. **Metrika** 70, 267–277.
[10] BOUZA-HERRERA, C., N., JUÁREZ-MORENO, P.O., SANTIAGO-MORENO, A. and SAUTTO-VALLEJO, J.M. (2022): A Two-Stage Scrambling Procedure: Simple and Stratified Random Sampling.to **Investigación Operacional** *; 43, 421-430.*
[11] CHAUDHURi, A., (1987): Randomized response surveys of finite populations: a unified approach with quantitative data. **J. Stat. Plann. Inference,** 15, 157–165.
[12] Chaudhuri, A. and Mukherjee, R. (1988): **Randomized Response: Theory and Techniques**. Marcel Dekker, New York.
[13] CHAUDHURI, A., CHRISTOFIDES, T.C. and RAO, C.R. (2016): **Data gathering, analysis and protection of privacy through randomized response techniques: qualitative and quantitative human traits**. North Holland. Amsterdam.
[14] CHAUDHURI, A., CHRISTOFIDES, T.C., SAHA, A., (2009): Protection of privacy in efficient application of randomized response techniques. **Stat. Methods Appl**. 18, 389–418.
[15] CHAUDHURY A. and H. STENGER (2005): **Sampling Survey**. Marcel Dekker, New York.
[16] EICHHORN, B., HAYRE, L.S.(1983) Scrambled randomized response methods for obtaining sensitive quantitative data. **J. Stat. Plann. Inference** 7, 307–316.
[17] ERIKSSON, S.A., (1973): A new model for randomized response. **Int. Stat. Rev**. 41, 40–43.
[18] GREENBERG, B. G., ABUL-ELA A.A., 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.
[19] GREENBERG, B. G., KUEBLER, R. R., JR., ABERNATHY, J. R., and HORVITZ, D. G. (1971): Application of the Randomized Response Technique in Obtaining Quantitative Data. **J Journal of the American Statistical Association**, 66, 334, 243-250
GUPTA, S., GUPTA, B., and SINGH, S. (2002): Estimation of sensitivity level of personal interview survey question. **Statistica,** 64, 643–653. https://doi.org/10.6092/issn.1973-2201/63
[20] GUPTA, S., THORNTON, B. (2002): Circumventing Social Desirability Response Bias in Personal Interview Surveys. **American Journal of Mathematical and Management Sciences**. 22~~,~~ 369-383.
[21] HORVITZ, D. G., SHAH, B. V., SIMMONS, WALT R. (1967): The unrelated question randomized response model. **Social Statistics Section Proceedings of the American Statistical Association**, 65-72.
[22] HUANG, K.C., (2008): Estimation of sensitive characteristics using optional randomized technique. **Qual. Quant.** 42, 679–686.
[23] HUANG, K. C. (2010): Unbiased estimators of mean, variance and sensitivity level for quantitative characteristics in finite population sampling. **Metrika** , 71:341–52.
[24] HUSSAIN, Z., (2012): Improvement of the Gupta and Thornton scrambling model through double use of randomization device, **International Journal Academic Research Business and Social Science**, 2, 91-97.
[25] HUSSAIN, Z. and SHAHID, M., I. (2019): Improved Randomized Response in Optional Scrambling Models. **DOI: https://doi.org/10.2991/jsta.d.191112.001.**
[26] KRUMPAL, I. (2012): Estimating the prevalence of xenophobia and anti-semitism in Germany: A comparison of the randomized response technique and direct questioning, **Social Science Research**. 41, 1387–1403.
[27] LARA, D., STRICKLER, J., DÍAZ, C. and ELLERTSON, C.. (2004): Measuring induced abortion in Mexico: A comparison of four methodologies, **Sociological Methods And Research**, 32 , 529–558.
[28] MANGAT, N.S., SINGH, S., (1994): An optional randomized response sampling technique. **J. Ind. Stat. Assoc**. 32, 71–75.

[29] MOORS, J.J.A.(1971): Optimization of the unrelated question randomized response model. **J. Amer. Statist. Assoc**., 66, 627-629.
[30] NATIONAL SURVEY ON THE DYNAMICS OF HOUSEHOLD RELATIONSHIPS (2016): **Instituto Nacional De Estadística Y Geografía.** https://bit.ly/3d4M43N. (last consulted 13 December, 2021.)
[31] PERRI,P.F., COBO, B. and RUEDA, M. (2017): A mixed-mode sensitive research on cannabis use and sexual addiction: Improving self-reporting by means of indirect questioning techniques, **Quality Quantity. DOI 10.1007/s11135-017-0537-0**
[32] PERRI, P.F, PELLE, E. and STRANGES, M. (2016): Estimating induced abortion and foreign irregular presence using the randomized response crossed model, Social Indicators Research 129, pp. 601–618.
[33] POLLOCK, K.H. and BEK, Y. (1976): A Comparison of Three Randomized Response Models for Quantitative Data, **Journal of the American Statistical Association,** 71, 884-886.
[34] RUEDA, M., COBO B., ARCOS, A. and ARNAB, R. (2016): Software for Randomized Response Techniques. In **Vol. 34 of Handbook of Statistics**, ed. A. Chaudhuri, T. C. Christofides and C. R. Rao, 155–164. Oxford, UK: Elsevier.
[35] RYU, J.B., KIM, J.-M., HEO, T.Y., PARK, C.G., (2006). On stratified randomized response sampling. **Model. Assist. Stat. Appl**. 1, 31–36
[36] SINGH, H., P. and GOREY, S. (2018): On Two Stage Optional Randomized Response Model. **Elixir Statistics** 123C (2018) 51963-51987.
[37] SALEEM, I., SANAULLAH, A. and KOYUNCU, N. (2019): Estimation of Mean of a Sensitive Quantitative Variable in Complex Survey: Improved Estimator and Scrambled Randomized Response Model. **GU J Sci 32, 1021-1043.** DOI: 10.35378/gujs.418681
[38] SINGH, S., JOARDER, A.H. (1997): Optional randomized response technique for sensitive quantitative variable. **Metron** 55, 151–157.
[39] STUBBE, J.H., CHORUS, A.M.J., FRANK, L.E, DE HON, O. and VAN DER HEIJDEN, P.G.M.. (2013): Prevalence of use of performance enhancing drugs by fitness center members, **Drug. Test. Anal**. 6, 434–438.
[40] WARNER, S.L. (1965): Randomized response: a survey technique for eliminating evasive answer bias. **Journal of the American Statistical Association**, 60, 63–69.