

RESEARCH ARTICLE

General Randomized Response Techniques Using Polya's Urn Process as a Randomization Device

Mashail M. Al-Sobhi ¹, Zawar Hussain ^{2*}, Bander Al-Zahrani ³

1. Department of Mathematics, Umm Alqura University 21421, Makkah, Saudi Arabia, **2.** Department of Statistics, Quaid-i-Azam University 45320, Islamabad, Pakistan, **3.** Department of Statistics, King Abdulaziz University 80203, Jeddah, Saudi Arabia

zhiangah@yahoo.com



CrossMark
click for updates

OPEN ACCESS

Citation: Al-Sobhi MM, Hussain Z, Al-Zahrani B (2014) General Randomized Response Techniques Using Polya's Urn Process as a Randomization Device. PLoS ONE 9(12): e115612. doi:10.1371/journal.pone.0115612

Editor: Enrique Hernandez-Lemus, National Institute of Genomic Medicine, Mexico

Received: April 12, 2014

Accepted: November 27, 2014

Published: December 26, 2014

Copyright: © 2014 Al-Sobhi et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper and its Supporting Information files.

Funding: The work has been funded by the Institute of Scientific Research and Revival of Islamic Heritage at Umm Alqura University (grant #43305030). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

Abstract

In this paper, interesting improvements in [1] and [2] randomized response techniques have been proposed. The proposed randomized response technique applies Polya's urn process (see [3]) to obtain data from respondents. One of the suggested technique requires reporting the number of draws to observe a fixed number of cards of certain type. On the contrary, the number of cards of a certain type is to be reported in case of second proposed randomized response model. Based on the information collected through the suggested techniques, two different unbiased estimators of proportion of a sensitive attribute have been suggested. A detailed comparative simulation study has also been done. The results are also supported by means of a small scale survey.

Introduction

Surveys and questionnaires are usual statistical tools for obtaining data about attitudes, behaviors, emotions, and so forth. The important assumption of any survey technique is that the respondents are completely truthful in their reporting. However, the legitimacy of this assumption is dubious when investigators ask questions that most would be hesitant to respond publicly. Examples of such questions are those that disclose whether the respondent possesses an illicit behavior, or a trait that is socially undesirable, or the question may concern with the trait of which respondent is embarrassed or which the respondent feels extremely personal to reveal openly. Faced with such questions, some respondents in a sample will decline to respond or will misreport. Either type of avoidance introduces a bias into collected information. Hence, there are serious procedural

hurdles to conduct surveys in studies in which a stigmatizing characteristic is associated with the phenomenon of concern.

To overcome these hurdles [4], developed the randomized response technique (RRT). [4] method is meant for estimating the proportion of a sensitive attribute prevailing in population of interest. It is based on the hypothesis that cooperation by the individuals should recover if their replies would not expose their true status. A number of variations of [4] RRT and new RRTs have been proposed by many researchers like [1], [2], [5], [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], [16], [17], and many others. In most of the RRTs, yielding either qualitative or quantitative response, reported response, commonly, follows either Bernoulli ([4], [18], [1], [9], [10], etc.), Poisson ([16]), multinomial ([8]), geometric ([2]) or negative hypergeometric distribution ([17]).

[1] introduced an ingenious RRT which produces the response following a Bernoulli distribution. A number of RRTs such as the [4], [10], [9] are special cases of [1] technique. [19] have reported that the [1]'s family of RRTs performs better than the Simmon's family in terms of efficiency and privacy protection. Recently [2], improved the [1]'s RRT by introducing geometric distribution as a randomization device.

There is an extensive amount of literature on the applications of urn models in different fields like genetics, capture-recapture models, computer theory, biology, learning processes, etc. Some of applications of urn models in epidemiology are noted by [20], [21] and [22], amongst many others. Applications to botany, lexicology and numismatics are mentioned by [23]. Interested readers may refer to [24] and [25] for a detailed account on the applications of urn models in randomization structures.

As far as RRT is concerned, urn scheme may be applied to determine the questions asked in a survey enquiry. The pioneer RRT proposed by [4], and further extended by [18] and [26] is a striking example. Utilization of urn model as a randomization device for [4] RRT may be briefly explained as follows. In the [4] RRT, the respondent randomly draws a card from an urn containing g green and r red cards. If a green card is drawn, the respondent will report (*yes* or *no*) to the question, "I am a member of sensitive group", and replace the card drawn. If a red card is drawn, the question is, "I am not a member of sensitive group". The interviewer is unaware of the colors of the cards drawn by the respondents, but the probability of drawing a green card is known. Under the assumptions of randomness of the drawing of balls and truthful reporting of answers, obviously, the total number of *yes* responses in a sample of n respondents follows a binomial distribution with parameters n and $g/(g+r)$.

Of the many urn models, the Polya's urn model is very popular within Statistics because it generalizes the binomial, hypergeometric, and beta-Bernoulli (beta-binomial) distributions through a single formula. In the present study, we intend to apply Polya's urn scheme to randomize the responses. It is important to note that different discrete distributions such as binomial, hypergeometric, negative binomial, geometric, negative hypergeometric, beta binomial, uniform, etc. can be generated through Polya's urn scheme. Thus, using Polya's urn schemes may be

taken as more flexible and a generalization of the above mentioned distributions. The idea is, actually, taken from the [2] RRT (a special case of [1] RRT) which yields a response following a geometric distribution. The rest of the paper is organized as follows. In the next section, we present the [1] and [2] RRTs. Two new estimators using Polya's urn process have been suggested in Section 3. Section 4 consists of discussion and conclusion of the study. A real life example has been presented in Section 5.

Some Recent Related RRTs

In this section, we present brief summaries of the [1] and [2] RRTs and introduce the notations. Let $U = (u_1, u_2, \dots, u_\infty)$ be an infinite dichotomous population and every individual in the population belongs either to a sensitive group (possessing a sensitive attribute) G , or to its complement \bar{G} . The problem is to estimate π ($0 < \pi < 1$), the unknown proportion of population members in group G . To do so, a sample $s = (u_1, u_2, \dots, u_n)$ of size n is drawn from the population U using a simple random sampling with replacement sampling scheme. Because of the sensitive nature of the attribute under study, a direct question regarding membership in G or otherwise is not expected to be helpful in terms of cooperation from the respondents. Thus, an alternative procedure such as RRT is needed if we are to procure reliable data on the sensitive attribute.

Two of the background RRTs are discussed in the following subsections.

1.1. Kuk [1] RRT

In this RRT, if a respondent belongs to a sensitive group G , then he/she is instructed to use a deck of cards having θ_1 proportion of cards with the statement, “ $I \in G$ ” and if he/she belongs to non-sensitive group \bar{G} , then he/she is requested to use a different deck of cards having θ_2 proportion of cards with the statement, “ $I \notin G$ ”. The probability of a *yes* response in the [1] model is given by

$$Pr(\text{yes})_{\text{kuk}} = \theta_{\text{kuk}} = \theta_1\pi + \theta_2(1 - \pi). \quad (2.1)$$

An unbiased estimator of π is given by

$$\hat{\pi}_{\text{kuk}} = \frac{\frac{n_1}{n} - \theta_2}{\frac{n}{(\theta_1 - \theta_2)}}, \quad \theta_1 \neq \theta_2, \quad (2.2)$$

where n_1 is the observed number of *yes* responses in the sample s and follows a binomial distribution with parameters $\theta_{\text{kuk}} = \theta_1\pi + \theta_2(1 - \pi)$ and n . Thus the variance of $\hat{\pi}_{\text{kuk}}$ is given by

$$V(\hat{\pi}_{\text{kuk}}) = \frac{\theta_{\text{kuk}}(1 - \theta_{\text{kuk}})}{n(\theta_1 - \theta_2)^2}. \quad (2.3)$$

1.2. Singh and Grewal [2] RRT

In this RRT, each respondent is provided with two decks of cards in the same way as in the [1] RRT. In the first deck of cards θ_1^* is the proportion of cards with the statement, “ $I \in G$ ” and $(1 - \theta_1^*)$ be the proportion of cards with the statement, “ $I \notin G$ ”. In the second deck of cards θ_2^* is the proportion of cards with the statement, “ $I \notin G$ ” and $(1 - \theta_2^*)$ be the proportion of cards with the statement, “ $I \in G$ ”. Up till here, it is same as that of the [1]. If a respondent belongs to sensitive group G , he/she is instructed to draw cards, one by one using with replacement, from the first deck of cards until he/she gets the first card bearing the statement of his/her own status, and requested to report the number of cards, say X , drawn by him/her to obtain the first card of his/her own status. If a respondent belongs to non-sensitive group \bar{G} , he/she is instructed to draw cards, one by one using with replacement drawing, from the second deck of cards until he/she gets the first card bearing the statement of his/her own status, and requested to report the number of cards, say Y , drawn by him/her to obtain the first card of his/her own status. Since cards are drawn using with replacement sampling, it is clear that X and Y follow geometric distributions with parameters θ_1^* and θ_2^* , respectively. If Z_i denotes the number of cards reported by the i th respondent, then it can be written as

$$Z_i = \alpha_i X_i + (1 - \alpha_i) Y_i,$$

where α_i is a Bernoulli random variable with $E(\alpha_i) = \pi$. The expectation of reported number of cards is given by

$$E(Z_i) = E(\alpha_i)E(X_i) + E(1 - \alpha_i)E(Y_i) = \frac{\pi}{\theta_1^*} + \frac{(1 - \pi)}{\theta_2^*}. \quad (2.4)$$

An unbiased estimator of π proposed by [2] is given by

$$\hat{\pi}_{SG} = \frac{\theta_1^* \theta_2^* \bar{Z} - \theta_1^*}{(\theta_2^* - \theta_1^*)}, \theta_1^* \neq \theta_2^*, \quad (2.5)$$

with variance given by

$$V(\hat{\pi}_{SG}) = \frac{\pi(1 - \pi)}{n} + \frac{\theta_2^{*2}(1 - \theta_1^*)\pi + \theta_1^{*2}(1 - \theta_2^*)(1 - \pi)}{n(\theta_2^* - \theta_1^*)^2}. \quad (2.6)$$

Proposed RRTs

In this section, we present two new RRTs using Polya's urn scheme.

3.1. Proposed RRT I

A more general RRT is explained below. Consider two decks having two types of cards, red and green. The deck 1 contains $a_1(b_1)$ red (green) cards. The deck 2 contains $a_2(b_2)$ red (green) cards. Each respondent belonging to the sensitive (non-sensitive) group is requested to use deck 1 (deck 2) and randomly draw $n_1(n_2)$ cards one by one. On each draw he/she is requested to replace the card drawn and add $c_1(c_2)$ cards of the same color. If a respondent belongs to sensitive (non-sensitive) group he is required to report the number of red cards drawn, say $X'(Y')$, in $n_1(n_2)$ draws. Obviously, here X' and Y' have the distributions $f_1(x')$ and $f_1(y')$, respectively. The functional forms of $f_1(x')$ and $f_1(y')$ are given by

$$f_1(x') = \binom{n_1}{x'} \frac{a_1^{(c_1, x')} b_1^{(c_1, n_1 - x')}}{(a_1 + b_1)^{(c_1, n_1)}}, x' = 0, 1, 2, \dots, n_1, \quad (3.1)$$

$$f_2(y') = \binom{n_2}{y'} \frac{a_2^{(c_2, y')} b_2^{(c_2, n_2 - y')}}{(a_2 + b_2)^{(c_2, n_2)}}, y' = 0, 1, 2, \dots, n_2, \quad (3.2)$$

where $r^{(s,j)} = r(r+s)(r+2s)\cdots[r+(j-1)s]$ for $r, s \in R^+$ and $j \in N$.

The response Z'_i from the i th respondent may be written as

$$Z'_i = \alpha_i X'_i + (1 - \alpha_i) Y'_i, \quad (3.3)$$

where α_i is a random variable defined as above and $E(X'_i) = \mu_{X'} = \frac{n_1 a_1}{a_1 + b_1}$ and $E(Y'_i) = \mu_{Y'} = \frac{n_2 a_2}{a_2 + b_2}$. Thus, expected response may be written as

$$E(Z'_i) = \pi \mu_{X'} + (1 - \pi) \mu_{Y'}. \quad (3.4)$$

Now an unbiased estimator of population proportion π may be defined and its variance can easily be worked out. By solving (3.4) for π and estimating $E(Z'_i)$ by $\bar{Z}' = (1/n) \sum_{i=1}^n Z'_i$, an unbiased estimator of π is suggested as follows:

$$\hat{\pi}_1 = \frac{\bar{Z}' - \mu_{Y'}}{\mu_{X'} - \mu_{Y'}}. \quad (3.5)$$

Its variance is given by

$$V(\hat{\pi}_1) = \frac{\pi(1 - \pi)}{n} + \frac{\pi \sigma_{X'}^2 + (1 - \pi) \sigma_{Y'}^2}{n(\mu_{X'} - \mu_{Y'})^2}, \quad (3.6)$$

where $\sigma_{X'}^2 = \frac{n_1 a_1 b_1}{(a_1 + b_1)^2} \left[1 + (n_1 - 1) \frac{c_1}{a_1 + b_1 + c_1} \right]$ and $\sigma_{Y'}^2 = \frac{n_2 a_2 b_2}{(a_2 + b_2)^2} \left[1 + (n_2 - 1) \frac{c_2}{a_2 + b_2 + c_2} \right]$.

Following remarks are in order.

Remark 1: It is interesting to see that the reported response Z'_i follows a two component mixture distribution with π and $(1-\pi)$ as the mixing probabilities. For $c_1=c_2=1$, the distribution of response Z'_i is a mixture of two beta-binomial distributions with parameters (n_1, a_1, b_1) and (n_2, a_2, b_2) .

Remark 2: For $c_1=c_2=0$, the distribution of Z'_i is a mixture of two binomial distributions with parameters $\left(n_1, p_1 = \frac{a_1}{a_1+b_1}\right)$ and $\left(n_2, p_2 = \frac{a_2}{a_2+b_2}\right)$.

Remark 3: For $c_1=c_2=-1$, the distribution of Z'_i is a mixture of two hypergeometric distributions with parameters (a_1+b_1, a_1, n_1) and (a_2+b_2, a_2, n_2) . In this case we must have $n_1 \leq a_1+b_1$ and $n_2 \leq a_2+b_2$.

Remark 4: If $a_1=b_1=c_1$ and $a_2=b_2=c_2$, the distribution of Z'_i is a mixture of two uniform distributions with parameters $(0, n_1)$ and $(0, n_2)$.

3.2. Proposed RRT II

The proposed RRT II works in a fashion similar to that of Proposed RRT I. Here, we assume that $c_1=c_2=1$, and respondents are requested to report the number of draws to observe a fixed number, say r_1 and r_2 , of red cards. Let $X''(Y'')$ denotes the number of draws from urn 1 (urn 2) required to observe $r_1(r_2)$ red cards. Obviously, now, X'' and Y'' have the distributions given by

$$g_1(x'') = \binom{x''-1}{r_1-1} \frac{a_1^{(1,r_1)} b_1^{(1,x''-r_1)}}{(a_1+b_1)^{(1,x'')}}, \quad x'' = r_1, r_1+1, r_1+2, \dots, \infty, \quad (3.7)$$

$$g_2(y'') = \binom{y''-1}{r_2-1} \frac{a_2^{(1,r_2)} b_2^{(1,y''-r_2)}}{(a_2+b_2)^{(1,y'')}}, \quad y'' = r_2, r_2+1, r_2+2, \dots, \infty, \quad (3.8)$$

The response Z''_i from the i th respondent may be written as $Z''_i = \alpha_i X''_i + (1-\alpha_i) Y''_i$, where α_i is a random variable defined as above and $E(X''_i) = \mu_{X''} = \frac{r_1(a_1+b_1-1)}{a_1-1}$ and $E(Y''_i) = \mu_{Y''} = \frac{r_2(a_2+b_2-1)}{a_2-1}$. Thus, expected response may be written as

$$E(Z''_i) = \pi \mu_{X''} + (1-\pi) \mu_{Y''}. \quad (3.9)$$

Now, following the steps as in subsection 3.1, an unbiased estimator of population proportion π may be defined and its variance can be derived. By solving (3.9) for π and estimating $E(Z''_i)$ by $\bar{Z}'' = n^{-1} \sum_{i=1}^n Z''_i$, an unbiased estimator of π is suggested as

$$\hat{\pi}_2 = \frac{\bar{Z}'' - \mu_{Y''}}{\mu_{X''} - \mu_{Y''}}. \quad (3.10)$$

Its variance is given by

$$V(\hat{\pi}_2) = \frac{\pi(1-\pi)}{n} + \frac{\pi\sigma_{X''}^2 + (1-\pi)\sigma_{Y''}^2}{n(\mu_{X''} - \mu_{Y''})^2}, \quad (3.11)$$

where

$$\begin{aligned} \sigma_{X''}^2 &= \frac{r_1(a_1 + b_1 - 1)}{(a_1 - 1)(a_1 - 2)} [b_1 + r_1(a_1 + b_1 - 2)] - r_1^2 \left(\frac{a_1 + b_1 - 1}{a_1 - 1} \right)^2 \\ \text{and } \sigma_{Y''}^2 &= \frac{r_2(a_2 + b_2 - 1)}{(a_2 - 1)(a_2 - 2)} [b_2 + r_2(a_2 + b_2 - 2)] \\ &\quad - r_2^2 \left(\frac{a_2 + b_2 - 1}{a_2 - 1} \right)^2 \end{aligned}$$

Remark 5: For $c_1 = c_2 = 0$, the distribution of Z''_i is a mixture of two negative binomial distributions with parameters $\left(\mathbf{r}_1, \mathbf{p}_1 = \frac{\mathbf{a}_1}{\mathbf{a}_1 + \mathbf{b}_1} \right)$ and $\left(\mathbf{r}_2, \mathbf{p}_2 = \frac{\mathbf{a}_2}{\mathbf{a}_2 + \mathbf{b}_2} \right)$.

Remark 6: For $c_1 = c_2 = -1$, the distribution of Z''_i is a mixture of two negative hypergeometric distributions with parameters $(a_1 + b_1, a_1, r_1)$ and $(a_2 + b_2, a_2, r_2)$.

Remark 7: For $r_1 = r_2 = 1$, the distribution of Z''_i is a mixture of two geometric distributions.

Remark 8: If $a_1 = b_1 = c_1$ and $a_2 = b_2 = c_2$, the distribution of Z''_i is a mixture of two uniform distributions.

Discussion and Conclusion

Since our objective in this study was to introduce an application of Polya's urn process to obtain data on sensitive variables, we did not intend to have a full-fledged comparative study of proposed estimators with any other estimators.

However, to have an idea, we just considered estimator $\hat{\pi}_2$ and compared it with [1] and [2] estimators assuming $c_1 = c_2 = 0$, $r_1 = r_2 = 4$, $\theta_1 = 0.7$ and $\theta_2 = 0.2$. The reason of setting $\theta_1 = 0.7$ and $\theta_2 = 0.2$ is that the [1] model is at its best when $|\theta_1 - \theta_2|$ is maximum. As mentioned in Remark 5 above, for $c_1 = c_2 = 0$, we have

$p_j = a_j(a_j + b_j)^{-1} = \theta_j^*$ for $j = 1, 2$. The relative efficiency (RE) of the estimator $\hat{\pi}_2$ relative to $\hat{\pi}_{Kuk}$ and $\hat{\pi}_{SG}$ is defined as $RE_1 = \frac{V(\hat{\pi}_{Kuk})}{V(\hat{\pi}_2)}$ and $RE_2 = \frac{V(\hat{\pi}_{SG})}{V(\hat{\pi}_2)}$,

respectively. The RE results are displayed in [S1 Table](#) available in the supporting information files. From [S1 Table](#) (see [S1 Table](#)), it observed that proposed estimator is relatively more efficient than that of [1] and [2]. For the situations,

where $c_1 = c_2 = 1$ and $c_1 = c_2 = -1$, we have done a bit detailed comparison study. As, n_1 (or n_2) is the fixed number of cards to be drawn in the proposed RRT I and r_1 (or r_2) is pre decided number of cards of certain type in proposed RRT II, we fixed $n_1 = r_1$ and $n_2 = r_2$ so that proposed estimators could be compared with each other on equal footings. The RE of the proposed estimators $\hat{\pi}_1$ and $\hat{\pi}_2$ relative to $\hat{\pi}_{Kuk}$ and $\hat{\pi}_{SG}$ is arranged in the [S2–S4 Tables](#) and [S5–S10 Tables](#), respectively. It was observed that RE of proposed estimators relative to $\hat{\pi}_{Kuk}$ increases with the increase in a_1 when $c_1 = c_2 = 1$ (see [S2](#) and [S3 Tables](#)). The proposed estimators are also more efficient when we take $c_1 = c_2 = -1$ (see [S4 Table](#)). Same is the behavior of RE of the proposed estimators when we compare them with $\hat{\pi}_{SG}$ (see [S5–S10 Tables](#)). From the [S2–S10 Tables](#) (see [S2–S10 Tables](#)), it is evident that the proposed estimators outshine the two competing estimators $\hat{\pi}_{Kuk}$ and $\hat{\pi}_{SG}$. Also, it can be observed that RE of both the estimators is higher (lower) for larger (smaller) π when either $c_1 = c_2 = 1$ or $c_1 = c_2 = -1$, whereas, for $c_1 = c_2 = 0$, the situation is reversed. The RE of proposed estimators is directly proportional to the difference between $n_1(r_1)$ and $n_2(r_2)$. The overall finding is that the proposed estimator $\hat{\pi}_1$ is comparatively more efficient than $\hat{\pi}_2$. That is, using number of cards of certain type in fixed drawings is more useful than forcing the respondent to keep drawing the cards until he/she observes a pre-decided number of cards of one kind.

It is to be noted that variances $V(\hat{\pi}_1)$ and $V(\hat{\pi}_2)$ are decreasing functions of $|\mu_{X'} - \mu_{Y'}|$ and $|\mu_{X''} - \mu_{Y''}|$, respectively. Thus, variances of the proposed estimators may be cut down to a desired level by suitably choosing the values of $a_1, a_2, b_1, b_2, n_1, n_2, r_1$ and r_2 so that $|\mu_{X'} - \mu_{Y'}|$ and $|\mu_{X''} - \mu_{Y''}|$ is a maximum.

Moreover, it is seen that the Polya's urn process generates different distributions. Thus, using Polya's urn process is more general and more flexible scheme to generate a randomized response following a desired distribution. Additionally, in the proposed RRTs, no additional sampling cost is needed and every respondent uses the same randomization device. These two features of the proposal may be considered as extra advantages associated with it.

A practical example

As a practical example, we conducted a small scale survey by drawing a sample of size 100. Consider the population of students currently enrolled in different programs at Quaid-i-Azam University, Islamabad. The students were requested verbally to volunteer themselves for this survey study and were assured that their identity will not be disclosed in anyway. From this population, we took 1000 students including 200 those students who had been using marijuana for the last six months. The purpose of this was to take a population with known population proportion of marijuana users, that is, we took $\pi = 0.2$. As from the simulation results, it is evident that the proposed RRT 1 is relatively better than the others, therefore, we decided to apply the proposed RRT 1 in actual application. A simple random sample of 100 students (out of 1000 selected students) was drawn using with replacement sampling and every selected student was given two urns each

containing red and green cards. The urn 1 (urn 2) contains 1000 (100) red and 300 (40) green cards. For generating data through proposed RRT 1, he/she, then, was asked to draw 3 (3) cards at random from urn 1 (urn 2) if he/she had used (not used) marijuana, at least once, in the last six months. At each draw, he/she was directed to replace two cards (i.e. $c_1 = c_2 = 1$) of the color of the card drawn. After drawing the cards, he/she was requested to report the number of red cards. For generating responses through [1] and [2] RRTs, we fixed $\theta_1 = 0.7, \theta_2 = 0.3$ and $\theta_1 = \theta_1^* = 0.7, \theta_2 = \theta_2^* = 0.3$. It is to be noted that the same respondents were taken to generate the responses through three different randomization devices considered in this study. The data obtained through these randomization devices are presented in [S1–S13 Tables](#) (see [S1–S13 Tables](#)). The estimates of the proportion of students who had used marijuana at least once, during the last six months, are obtained as $\hat{\pi}_1 = 0.204$, $\hat{\pi}_{Kuk} = 0.12$ and $\hat{\pi}_{SG} = 0.2475$. From these estimates, it is clear that the proposed RRT 1 provided the closest estimate of the population proportion, i.e. $\pi = 0.2$. Hence, the proposed RRT 1 is more accurate than the other RRTs considered in this small scale survey.

Supporting Information

S1 Table. RE_1 and RE_2 values for $\theta_1 = 0.7, \theta_2 = 0.2$ and $0.1 \leq \pi \leq 0.9$.

[doi:10.1371/journal.pone.0115612.s001](#) (DOCX)

S2 Table. Relative efficiency of $\hat{\pi}_1$ (**in bold**) $\hat{\pi}_2$ with respect to $\hat{\pi}_{Kuk}$ for $\theta_1 = 0.7, \theta_2 = 0.3, n_2 = r_2 = 6, c_1 = c_2 = 1, a_1 = 9, b_1 = 3, a_2 = 10, b_2 = 4$.

[doi:10.1371/journal.pone.0115612.s002](#) (DOCX)

S3 Table. Relative efficiency of $\hat{\pi}_1$ (**in bold**) $\hat{\pi}_2$ with respect to $\hat{\pi}_{Kuk}$ $\theta_1 = 0.7, \theta_2 = 0.3, n_2 = r_2 = 6, c_1 = c_2 = 1, a_1 = 11, b_1 = 3, a_2 = 10, b_2 = 4$,

[doi:10.1371/journal.pone.0115612.s003](#) (DOCX)

S4 Table. Relative efficiency of $\hat{\pi}_1$ (**in bold**) $\hat{\pi}_2$ with respect to $\hat{\pi}_{Kuk}$ $\theta_1 = 0.7, \theta_2 = 0.3, n_2 = r_2 = 6, c_1 = c_2 = -1, a_1 = 9, b_1 = 3, a_2 = 10, b_2 = 4$,

[doi:10.1371/journal.pone.0115612.s004](#) (DOC)

S5 Table. Relative efficiency of $\hat{\pi}_1$ (**in bold**) $\hat{\pi}_2$ with respect to $\hat{\pi}_{SG}$ for $n_1 = r_1 = 1, n_2 = r_2 = 6, c_1 = c_2 = 1, a_1 = 9, b_1 = 3, a_2 = 10, b_2 = 4, \theta_1^* = 0.1$.

[doi:10.1371/journal.pone.0115612.s005](#) (DOCX)

S6 Table. Relative efficiency of $\hat{\pi}_1$ (**in bold**) $\hat{\pi}_2$ with respect to $\hat{\pi}_{SG}$ for $n_1 = r_1 = 1, n_2 = r_2 = 6, c_1 = c_2 = 1, a_1 = 9, b_1 = 3, a_2 = 10, b_2 = 4, \theta_1^* = 0.3$.

[doi:10.1371/journal.pone.0115612.s006](#) (DOCX)

S7 Table. Relative efficiency of $\hat{\pi}_1$ (**in bold**) $\hat{\pi}_2$ with respect to $\hat{\pi}_{SG}$ for $n_1 = r_1 = 1, n_2 = r_2 = 6, c_1 = c_2 = 1, a_1 = 9, b_1 = 3, a_2 = 10, b_2 = 4, \theta_1^* = 0.5$.

[doi:10.1371/journal.pone.0115612.s007](#) (DOCX)

S8 Table. Relative efficiency of $\hat{\pi}_1$ (**in bold**) $\hat{\pi}_2$ with respect to $\hat{\pi}_{SG}$ for $n_1 = r_1 = 1, n_2 = r_2 = 6, c_1 = c_2 = 1, a_1 = 9, b_1 = 3, a_2 = 10, b_2 = 4, \theta_1^* = 0.7$.

[doi:10.1371/journal.pone.0115612.s008](#) (DOCX)

S9 Table. Relative efficiency of $\hat{\pi}_1$ (**in bold**) $\hat{\pi}_2$ with respect to $\hat{\pi}_{SG}$ for $n_1=r_1=1$, $n_2=r_2=6$, $c_1=c_2=1$, $a_1=9$, $b_1=3$, $a_2=10$, $b_2=4$, $\theta_1^*=0.9$.

[doi:10.1371/journal.pone.0115612.s009](https://doi.org/10.1371/journal.pone.0115612.s009) (DOCX)

S10 Table. Relative efficiency of $\hat{\pi}_1$ (**in bold**) $\hat{\pi}_2$ with respect to $\hat{\pi}_{SG}$ for $n_1=r_1=1$, $n_2=r_2=6$, $c_1=c_2=-1$, $a_1=9$, $b_1=3$, $a_2=10$, $b_2=4$, $\theta_1^*=0.1$.

[doi:10.1371/journal.pone.0115612.s010](https://doi.org/10.1371/journal.pone.0115612.s010) (DOCX)

S11 Table. Data obtained through proposed RRT 1 using $n_2=n_2=3$, $c_1=c_2=1$, $a_1=1000$, $b_1=30$, $a_2=100$, $b_2=40$.

[doi:10.1371/journal.pone.0115612.s011](https://doi.org/10.1371/journal.pone.0115612.s011) (DOC)

S12 Table. Data obtained through [1] RRT using $\theta_1=0.7$, $\theta_2=0.3$.

[doi:10.1371/journal.pone.0115612.s012](https://doi.org/10.1371/journal.pone.0115612.s012) (DOCX)

S13 Table. Data obtained through [2] RRT using $\theta_1^*=0.3$, $\theta_2^*=0.7$.

[doi:10.1371/journal.pone.0115612.s013](https://doi.org/10.1371/journal.pone.0115612.s013) (DOCX)

Author Contributions

Conceived and designed the experiments: MA ZH BA. Performed the experiments: ZH. Analyzed the data: ZH. Contributed reagents/materials/analysis tools: ZH. Wrote the paper: ZH MA BA.

References

1. Kuk AYC (1990) Asking Sensitive Questions Indirectly. *Biometrika* 77: 436–438.
2. Singh S, Grewal IS (2013) Geometric distribution as a randomization device: implemented to the Kuk's model. *International Journal of Contemporary Mathematical Sciences* 8: 243–248.
3. Mahmoud H (2008) Pólya urn models: CRC press.
4. Warner SL (1965) Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of the American Statistical Association* 60: 63–69.
5. Horvitz DG, Shah BV, Simmons WR (1967) The unrelated question randomized response model. *Proceedings of the Social Statistics Section, American Statistical Association*: 65–72.
6. Moors JJA (1971) Optimization of the Unrelated Question Randomized Response Model. *Journal of the American Statistical Association* 66: 627–629.
7. Folsom RE, Greenberg BG, Horvitz DG, Abernathy JR (1973) The Two Alternate Questions Randomized Response Model for Human Surveys. *Journal of the American Statistical Association* 68: 525–530.
8. Kim J, Flueck JA (1978) An additive randomised response model. In proceedings of the Survey Research Section, American Statistical Association: 351–355.
9. Mangat NS, Singh R (1990) An Alternative Randomized Response Procedure. *Biometrika* 77: 439–442.
10. Mangat NS (1994) An Improved Randomized Response Strategy. *Journal of the Royal Statistical Society Series B (Methodological)* 56: 93–95.
11. Christofides TC (2003) A generalized randomized response technique. *Metrika* 57: 195–200.
12. Chang H-J, Huang K-C (2001) Estimation of proportion and sensitivity of a qualitative character. *Metrika* 53: 269–280.
13. Gjestvang CR, Singh S (2006) A New Randomized Response Model. *Journal of the Royal Statistical Society Series B (Statistical Methodology)* 68: 523–530.

14. Perri PF (2008) Modified randomized devices for Simmons' model. *Model Assisted Statistics and Applications* 3: 233–239.
15. Odumade O, Singh S (2009) Improved Bar-Lev, Bobovitch, and Boukai Randomized Response Models. *Communications in Statistics - Simulation and Computation* 38: 473–502.
16. Gi-Sung Lee K-HH, Chang-Kyo Son, Jong-Min Kim (2012) An Extension of Quaternber's Randomized Response Model. *Journal of the Korean Data Analysis Society* 15: 1755–1766.
17. Singh S, Sedory S (2013) A new randomized response device for sensitive characteristics: an application of the negative hypergeometric distribution. *METRON* 71: 3–8.
18. Greenberg BG, Abul-Ela A-LA, Simmons WR, Horvitz DG (1969) The Unrelated Question Randomized Response Model: Theoretical Framework. *Journal of the American Statistical Association* 64: 520–539.
19. Guerrero M, Sandri MF (2007) A note on the comparison of some randomized response procedures. *Journal of Statistical Planning and Inference* 137: 2184–2190.
20. Weiss GH (1965) On the Spread of Epidemics by Carriers. *Biometrics* 21: 481–490.
21. Dietz K (1966) On the Model of Weiss for the Spread of Epidemics by Carriers. *Journal of Applied Probability* 3: 375–382.
22. Kryscio RJ (1974) On the Extended Simple Stochastic Epidemic Model. *Biometrika* 61: 200–202.
23. Faucounau J (1975) Note sur une loi de probabilité discrète mkconnue. *Math et Sci Humaines* 52: 55–68.
24. Johnson NL, Kotz S (1977) Urn models and their application: an approach to modern discrete probability theory: Wiley New York.
25. Rosenberger WF (2002) Randomized urn models and sequential design. *Sequential Analysis* 21: 1–28.
26. Liu PT, Chow LP (1976) A New Discrete Quantitative Randomized Response Model. *Journal of the American Statistical Association* 71: 72–73.