https://doi.org/10.56093/JISAS.V80I1.17
Available online at https://isas.live/newsite
JOURNAL OF THE INDIAN SOCIETY OF AGRICULTURAL STATISTICS 80(1) 2026 187–206

Randomized Response Techniques: Its Development in Six Decades

Raghunath Arnab

University of Kwazulu-Natal, Durban, South Africa

Received 07 February 2026; Revised 08 April 2026; Accepted 21 April 2026

SUMMARY

Randomized response techniques (RR) are used to collect data relating to sensitive, personal, and /or stigmatizing characteristics. The direct method (DR) of interview cannot be applied in such situations because the respondents generally do not reveal the true information due to the social stigma and /or fear. So, the respondents often provide socially approved or untrue responses. As a result, DR suffers from inaccurate data and a high rate of nonresponse. To overcome such difficulties, Warner (1965) introduced an ingenious method of data collection, known as the randomized response (RR) technique, where respondents need not provide the true response but only to provide the outcome of a random experiment. Respondents' privacy is maintained, as the interviewer will not identify whether the submitted response is true or masked. A chronological development of the RR techniques for the last six decades is provided, starting with the pioneering RR model provided by Warner. Detailed developments of the RR models for the qualitative and quantitative characteristics are provided along with the applications for the complex survey designs. Methods of estimation of the population characteristics from the multiple randomized response techniques are presented. The methods of estimations of the proportions of overlaps of multiple characteristics are presented. The optional randomized response techniques (ORT) are also discussed. Finally, some of the nonrandomized response techniques (NRR) and the measures of privacy protection are presented.

Keywords: Crosswise model; Complex survey; Direct response; Forced RR Model; Maximum likelihood; Multiple randomized response, Non-randomized response; Parallel model; Triangular Model; Primary protection.

1. INTRODUCTION

In various surveys, information related to highly sensitive, stigmatized, and personal questions, such as domestic violence, drug addiction, sexual habits, and political affiliations, among others, is collected. In such situations, respondents very often provide untrue or socially approved responses or even refuse to respond because of social stigma or fear that their personal information may be disclosed to the others. Warner (1965) provided an indirect method of interview, which is known as the "randomized response (RR) technique," where the respondent need not answer directly the questions but answers some questions related to the outcome of a certain randomization technique. So, in this procedure, respondents' privacy is protected, and the response rate is enhanced. Warner's technique was extended by several authors, including Horvitz *et al.* (1967), Greenberg *et al.* (1969), Boruch (1972), Takahasi and Sakasegawe (1977), Franklin (1989), Arnab (1990, 1996), Kuk (1990), Mangat and Singh (1990), Singh and Singh (1992, 1993), and Arnab and Rueda (2016), among others.

RR techniques in real surveys were conducted by various researchers, e.g., Abernathy *et al.* (1970) related to the incidence of induced abortions; Folsom *et al.* (1973) regarding drinking and driving; and Goodstadt and Gruson (1975) concerning drug use. Lara *et al.* (2004) used the unrelated question randomized response technique in estimating the proportion of induced abortion attempts among women in Mexico. Srivastava *et al.* (2015) concerning sexual abuse of children aged 11–17 years in Uttar Pradesh, India, in 2011. Cobo *et al.* (2016) concerning consumption of cannabis cigarettes among Spanish university students. Mieth *et al.* (2021) concerning compliance with COVID regulations in Germany. Mielecka-Kubień

E-mail address: arnab.raghunah@gmail.com

and Wójcik (2022) concerning the prevalence of illicit drug use among high school students in the Silesian voivodeship (Poland).

A few researchers, including Geurts (1980), Dalton & Metzger (1992), Hubbard, Casper & Lessler (1989), Tian *et al.* (2009), Rueda *et al.* (2018), and Arnab *et al.* (2019), pointed out some demerits of the RR techniques, such as the interviewers needing to explain the techniques clearly to the respondents and the respondents often misunderstanding the logic behind the RR techniques. Hence, the RR technique becomes more expensive and time-consuming than the DR and not applicable to the mail inquiry. Respondents often provide different responses if they repeat the experiment again. Furthermore, the RR method possesses lower efficiency than the DR. Comprehensive reviews of RR techniques are obtained in Fox and Tracy (1986), Chaudhuri and Mukerjee (1988), Arnab (2017), and Arnab (2025), among others.

2. THREE BASIC RR TECHNIQUES

In this section, we will describe the pioneering RR technique that was introduced by Warner (1965) and its important developments. Consider a finite population $U=(1,...,i,...,N)$ of N units that is classified into two categories A and its complement $\bar{A}$. The category A comprises those persons who possess certain sensitive attributes, A such as HIV +, and others $\bar{A}$ that are complementary, which may be sensitive or may not be sensitive. The proportion of persons belonging to the sensitive group of the population U is π which is unknown. For estimation of π, a sample s of n units is selected from U by the simple random sampling with replacement (SRSWR) method. Since the subject of inquiry is sensitive, selected respondents were not asked about their membership in the group A or its complement $\bar{A}$, but each of them was asked to perform a randomized response trial in the absence of the interviewer and requested to report the outcome of the trial only. The interviewer will only know the reported outcome of the randomized response. Hence, the privacy of the true status of the respondents (membership of A or $\bar{A}$) is protected.

2.1 Warner method (W-method)

In the Warner's (1965) method, respondents are requested to draw one card, unnoticed by the interviewer, at random from a well-shuffled pack containing two types of cards identical in appearance. Type 1, with proportion $P(\neq 1/2)$, bears the question "Q1: Do you belong to the sensitive group A?" and the Type 2, bears the question "Q2: Do you belong to the sensitive group $\bar{A}$?" Respondents should report truthfully "yes" or "no" as his or her response. Privacy is protected, as the interviewer will know the answer "yes" or "no" but will not identify which of the questions, Q1 or Q2, was answered.

$$\overbrace{\begin{matrix} Q1: Do \text{ you belong to the group } A? & Q2: Do \text{ you belong to the group } \bar{A}? \\ (\textit{with} \text{ probability P}) & (\textit{with} \text{ probability } 1-\text{P}) \end{matrix}}^{\textit{Answer} \text{ truthfully the following questions:}}$$

Fig. 2.1. W-method

Estimation of π :

Let θ_W be the probability of obtaining a "yes" answer from a respondent selected at random. Then,

$$\begin{aligned}\theta_W &= \text{Prob (obtaining a "yes" answer } \cap \textit{ respondent} \in \text{A)} + \text{Prob (obtaining a "yes" answer } \cap \textit{respondent} \in \bar{\text{A}}) \\ &= \text{Prob (obtaining a "yes" answer} | \textit{ respondent} \in \text{A)} \times \text{Prob (}\textit{respondent} \in \text{A)} + \text{Prob (obtaining a "yes" answer} | \textit{ respondent} \in \bar{\text{A}}) \times \text{Prob (}\textit{respondent} \in \bar{\text{A}}) \\ &= P\pi + (1-P)(1-\pi) \\ &= \pi(2P-1) + (1-P)\end{aligned} \tag{2.1.1}$$

Further, let $\hat{\theta}_W$ be the proportion of respondents that answered "yes" in the sample. Then $\hat{\theta}_W$ follows a $B(n,\pi)$ binomial distribution with the parameter π. Then one finds

(i) $E(\hat{\theta}_W) = n\theta_W$, (ii) $Var(\hat{\theta}_W) = \dfrac{\theta_W(1-\theta_W)}{n}$ and

(iii) an unbiased estimator of $Var(\hat{\theta}_W)$ is

$$V\hat{a}r(\hat{\theta}_W) = \frac{\hat{\theta}_W(1-\hat{\theta}_W)}{n-1} \tag{2.1.2}$$

Eq. (2.1.2) yields

Theorem 2.1.1.

(i) $\hat{\pi}_{\mathrm{W}} = \dfrac{\hat{\theta}_{\mathrm{W}} - (1-P)}{2P-1}$,

$P \neq 1/2$ is an unbiased estimator of π

(ii) The variance of $\hat{\pi}_{\mathrm{W}}$ is

$$\mathrm{Var}(\hat{\pi}_{\mathrm{W}}) = \frac{\pi(1-\pi)}{n} + \frac{P(1-P)}{n(2P-1)^2}$$

(iii) An unbiased estimator of $\mathrm{Var}(\hat{\pi}_{\mathrm{W}})$ is

$$\hat{\mathrm{Var}}(\hat{\pi}_{\mathrm{W}}) = \frac{\hat{\theta}_{\mathrm{W}}(1-\hat{\theta}_{\mathrm{W}})}{(n-1)(2P-1)^2}$$

Warner estimator $\hat{\theta}_W$ is less efficient than $\hat{\pi}$, the sample proportion of "yes" answers as an estimator for estimating the population proportion π. The relative efficiency of the Warners estimator is

$$E_W = \frac{\mathrm{Var}(\hat{\pi})}{\mathrm{Var}(\hat{\pi}_{\mathrm{W}})}$$

$$= \frac{\pi(1-\pi)}{n}\left[\frac{\pi(1-\pi)}{n} + \frac{P(1-P)}{n(2P-1)^2}\right]^{-1} < 1 \qquad (2.1.3)$$

Singh (1976, 1978) pointed out that $\hat{\pi}_{\mathrm{W}}$ proposed by Warner is not the maximum likelihood estimator (MLE) as it can take values outside of $[0,1]$, the parameter space of π. The MLE of π is

$$\hat{\pi}_{\mathrm{MLE}} = \begin{cases} \hat{\pi}_{\mathrm{W}}, & \text{if } 0 \leq \hat{\pi}_{\mathrm{W}} \leq 1 \\ 1, & \text{if } \hat{\pi}_{\mathrm{W}} > 1 \\ 0, & \text{if } \hat{\pi}_{\mathrm{W}} < 0 \end{cases} \qquad (2.1.4)$$

2.2 Unrelated question model (U-model)

In Warner's RR model, respondents are asked to answer whether he or she belongs to the sensitive group A or its complementary group $\overline{A}$. The group $\overline{A}$ might be sensitive or not. In this case, the respondent may not feel totally comfortable answering the questions. To overcome such difficulties, Horvitz *et al.* (1967) and Greenberg *et al.* (1969) proposed an alternative model, known as the unrelated question model (U-model) or Simmon's model. The proposed model is similar to the W-model, but the question $Q2$ was replaced by the question "Q_B: Do you belong to the group B?" The group B is totally unrelated to the character A or completely innocuous such as a Christian or a vegetarian. The proportion of persons belonging to group B in the population π_B is known. So, in the U-method, respondents were asked to draw one card at random from a pack containing two types of cards $Q1$ and Q_B, with proportions P and $1-P$ as in the W-model, and requested to answer "yes" or "no" truthfully to the question written on the card drawn.

$$\overbrace{\begin{array}{cc} Q1: \textit{Do} \text{ you belong to the} & QB: \textit{Do} \text{ you belong to the} \\ \text{group } A? & \text{group } B? \\ (\textit{with} \text{ probability P}) & (\textit{with} \text{ probability } 1-\text{P}) \end{array}}^{\textit{Answer} \text{ truthfully the following questions:}}$$

Fig. 2.2. U-model

2.2.1. Estimation of π

For the U-model, the probability of getting a "yes" answer from a respondent selected at random is

$$\theta_{\mathrm{U}} = \mathrm{Prob}(Yes \mid \text{respondent} \in \mathrm{A}) \times \mathrm{Prob}(\text{respondent} \in \mathrm{A}) + \mathrm{Prob}(Yes \mid \text{respondent} \in \mathrm{B}) \times \mathrm{Prob}(\text{respondent} \in \mathrm{B})$$

$$= \pi P + \pi_B(1-P) \qquad (2.2.1)$$

Let $\hat{\theta}_{\mathrm{U}}$ be the sample proportion of "yes" answers. Then,

Theorem 2.2.1.

(i) $\hat{\pi}_{\mathrm{U}} = \dfrac{\hat{\theta}_{\mathrm{U}} - (1-P)\pi_B}{P}$ is an unbiased estimator of π

(ii) The variance of $\hat{\pi}_{\mathrm{U}}$ is

$$\mathrm{Var}(\hat{\pi}_{\mathrm{U}}) = \frac{\theta_{\mathrm{U}}(1-\theta_{\mathrm{U}})}{nP^2}$$

(iii) An unbiased estimator of $\mathrm{Var}(\hat{\pi}_{\mathrm{U}})$ is

$$\hat{\mathrm{Var}}(\hat{\pi}_{\mathrm{U}}) = \frac{\hat{\theta}_{\mathrm{U}}(1-\hat{\theta}_{\mathrm{U}})}{(n-1)P^2}$$

In case π_{B} is unknown, π is estimated from two independent samples s_1 and s_2 of sizes n_1 and $n_2(=n-n_1)$. Details are given in Chaudhuri and Mukerjee (1988), and Arnab (2025).

2.3 Forced RR model (F-model)

Boruch (1972) and Horvitz *et al.* (1976) proposed the F-model. In this-model, the respondents are requested to draw a marble at random from a bag containing three types of marbles: red, yellow, and green with proportions P, α and $\beta = 1 - P - \alpha$ $(0 < P, \alpha, \beta < 1)$. If a red marble is drawn, the respondent should answer "yes" or "no" to the sensitive question "Q1: Do you belong to the sensitive group A ?" If a yellow marble is selected, the respondent should answer "yes" only. If a green marble is selected, the respondent should answer "no" only.

Directions:

$\downarrow$

Answer "yes" or "no" Do you belong to the group A? (with probability P)	Answer "yes" (*with probability* α)	Answer "no" *with probability* $(1 - P - \alpha)$

Fig 2.3.1. F-model

Fig. 2.3.1. F-model

For the F-model probability of obtaining a "yes" answer is

$$\theta_{\mathrm{F}} = \pi P + \alpha \tag{2.3.1}$$

Let $\hat{\theta}_{\mathrm{F}}$ be the proportion of "yes" answers. Then, the following theorem is obtained:

Theorem 2.3.1.

(i) $\hat{\pi}_{\mathrm{F}} = \dfrac{\hat{\theta}_{\mathrm{F}} - \alpha}{P}$ is an unbiased estimator of π

(ii) The variance of $\hat{\pi}_{\mathrm{F}}$ is

$$\mathrm{Var}(\hat{\pi}_{\mathrm{F}}) = \frac{\theta_{\mathrm{F}}(1 - \theta_{\mathrm{F}})}{nP^2}$$

(iii) An unbiased estimator of $\mathrm{Var}(\hat{\pi}_{\mathrm{F}})$ is

$$\hat{\mathrm{Var}}(\hat{\pi}_{\mathrm{F}}) = \frac{\hat{\theta}_{\mathrm{F}}(1 - \hat{\theta}_{\mathrm{F}})}{(n-1)P^2}$$

Several alternative RR techniques that include Mangat and Singh (1990), Mangat (1992), Mangat, Singh, and Singh (1992), Singh (1993), Chang-Liang (1996), Singh and Joarder (1997), Singh, Horn, Singh, and Mangat (2003), Chang, Wang, and Huang (2004), and Singh and Mathur (2003), among others, are proposed in the literature. Most of them are combinations of the basic three models mentioned above.

3. RR TECHNIQUE FOR QUANTITATIVE VARIABLES

The RR technique for the quantitative sensitive characteristics, such as number of abortions over a life, amount of illegal income, expenditure on gambling, numbers of sextual partners, etc., was first proposed by Greenberg *et al.* (1971). The method was extended by Eriksson (1973). Poole (1974), Duffy and Waterton (1978), Pollock and Beck (1976), and Eichhorn and Hayre (1983), among others. Combining the additive model and the multiplicative model, the distributive model is proposed by Arnab (2025). The most general RR model was proposed by Arnab (1990), which can be applied for both the qualitative and quantitative characteristics.

3.1 Greenberg *et al.* (1971) RR technique

In this technique, a sample n of respondents is selected by the SRSWR method. Each of the selected respondents is asked to answer one of the following two questions using a suitable randomized device.

"Q1 How much did the household's head earn (y) *last year?"* with a probability P

and

"Q2: How much do you think the average household head earns in a year (x) *?"* with a probability $1 - P$.

Let z be the RR provided by a respondent. Then,

(i) $E(z) = P\mu_y + (1-P)\mu_x$ and

(ii) $E(z^2) = P(\mu_y^2 + \sigma_y^2) + (1-P)(\mu_x^2 + \sigma_x^2)$ (3.1.1)

where

$E(y) = \mu_y, E(x) = \mu_x, V(y) = \sigma_y^2$ and $V(x) = \sigma_x^2$

Eq. (3.1.1) yields the following theorem:

Theorem 3.1.1.

Let $z_1, \ldots, z_i, \ldots, z_n$ be the RR obtained from a sample of n respondents, then

(i) $\hat{\mu}_y = \dfrac{\bar{z} - (1-P)\mu_x}{P}$ is an unbiased estimator of μ_y where $\bar{z} = \sum_{i=1}^{n} z_1 / \mathrm{n}$.

(ii) The variance of $\hat{\mu}_y$ is

$$\mathrm{Var}(\hat{\mu}_{yG}) = \frac{\sigma_z^2}{nP^2}, \sigma_z^2$$
$$= P\sigma_y^2 + (1-P)\sigma_x^2 + P(1-P)(\mu_y - \mu_x)^2$$

(iii) An unbiased estimator of $\mathrm{Var}(\hat{\mu}_{yG})$ is

$$\hat{\mathrm{Var}}(\hat{\mu}_{yG}) = \frac{s_z^2}{nP^2}, s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})^2.$$

3.2 Eriksson's RR technique

In Eriksson's (1973) RR model, each of the selected respondents in the sample is asked to report the true response y with a known probability $c(0<c<1)$ or a scramble response $Q_j\,(j=1,..,M)$ with a probability $q_j\;(0<q_j<1, \sum_{j=1}^{M} q_j = 1-c)$. The vector $\mathbf{Q} = (Q_1,\cdots,Q_M)$ is known and expected to cover the range of y values. Let z_i be the randomized response obtained from the i th $(i=1,..,N)$ respondent. Then

$$z_i = \begin{cases} y_i, \text{ with probability } c \\ Q_i, \text{ with probability } q_i \end{cases} \tag{3.2.1}$$

Denoting by E_R and V_R expectation and variance respect to the RR technique, one finds

$$E_R(z_i) = cy_i + (1-c)\mu_Q \text{ and}$$
$$V_R(z_i) = (1-c)\left[c(y_i - \mu_Q)^2 + \sigma_Q^2\right] \tag{3.2.2}$$

where $\mu_Q = \sum_{j=1}^{M} q_j^* Q_j$, $q_j^* = q_j/(1-c)$ and

$\sigma_Q^2 = (\sum_{j=1}^{M} q_j^* Q_j^2 - \mu_Q^2)$.

From the Eq. (3.3.2) one obtains an unbiased estimator of y_i as

$$r_i = \left\{z_i - (1-c)\mu_Q\right\}/c \tag{3.2.3}$$

Finally, Eq. (3.2.3) yields

Theorem 3.2.1

Let $z_1, z_2, ..., z_n$ be the RR obtained from a sample of n respondents selected by the SRSWR method. Then,

(i) $\bar{r} = \frac{1}{n}\sum_s r_i$ is an unbiased estimator of the population mean $\bar{Y} = \frac{1}{N}\sum_{i\in U} y_i$

where $\sum_s$ and $\sum_{i\in U}$ denote respectively the sum of the units in s including repetition and distinct units in U.

(ii) The variance of $\bar{r}$ is

$$\mathrm{Var}(\bar{r}) = \frac{1}{n}(\sigma_y^2 + \bar{\phi})$$

where $\sigma_y^2 = \frac{1}{N}\sum_{i\in U}(y_i - \bar{Y})^2$, $\bar{Y} = \sum_{i\in U} y_i / N$ and $\bar{\phi} = \frac{1}{N}\sum_{i\in U}\phi_i$

(iii) An unbiased estimator of $\mathrm{Var}(\bar{r})$ is

$$\hat{\mathrm{Var}}(\bar{r}) = \frac{1}{n(n-1)}\sum_s (r_i - \bar{r})^2$$

Most of the methods available in the literature may be regarded as a particular case of Eriksson's (1973) model. Some of the methods are described as follows:

3.2.1. Additive model

Poole (1974), and Duffy and Waterton (1984) proposed the additive model. In this model,

$$z_i = y_i + x_i \tag{3.2.4}$$

where x_i is a random sample from a population with a known mean μ_x and known variance σ_x^2

3.2.2. Multiplicative model

Pollock and Bek (1976), and Eichhorn and Hayre (1983) proposed the multiplicative model with

$$z_i = y_i x_i \tag{3.2.5}$$

3.2.3. Distributive model

Proposed by Arnab (2025) with

$$z_i = x_{1i} y_i + x_{2i} \tag{3.2.6}$$

where x_{1i} and x_{2i} random sample from populations with known means μ_{x1}, μ_{x2} and known variances $\sigma_{x1}^2, \sigma_{x2}^2$ respectively.

3.2.4. Bar-lev et al.'s model

$$z_i = \begin{cases} y_i, & \text{with probability } p \\ y_i x_i, & \text{with probability } 1-p \end{cases} \quad (3.2.7)$$

where x_i is a random sample from a distribution with known mean μ_x and variance σ_x^2, such as normal or chi-square.

3.3. Arnab's RR model

Arnab (1990) proposed the most general RR model, which can be applied for both the qualitative and quantitative variables. Most of the RR models become a particular case of the proposed model. Let z_i be the RR obtained from the ith respondent, and $r_i = ø(z_i)$ be a function of z_i such that the transformed response $r_i = ø(z_i)$ satisfies the model

$$E_R(r_i) = y_i,\ V_R(r_i) = \phi_i \text{ and } C_R(r_i, r_j) = 0 \text{ for } i \neq j \quad (3.3.1)$$

where $\phi_i = \phi_i(y_i)$ is a function of y_i, and the responses z_i and z_j are independent for $i \neq j$.

In the case $\phi_i = \alpha y_i^2 + \beta y_i + \gamma$, an unbiased estimator ϕ_i is

$$\hat{\phi}_i = \frac{1}{\alpha+1}\{\alpha r_i^2 + \beta r_i + \gamma\} \quad (3.3.2)$$

We will show the following RR models follows Arnab's (1990) model.

3.3.1. Qualitative variables

For the qualitative variable y, we will define

$$y_i = \begin{cases} 1, & if\ i \in A \\ 0, & if\ i \notin A \end{cases} \text{ and}$$

$$z_i = \begin{cases} 1, & \text{if the respondent answer "yes"} \\ 0, & \text{if the respondent answer "no"} \end{cases} \quad (3.3.3)$$

Warner's (1965) model:

$$E_R(z_i) = Py_i + (1-P)(1-y_i)$$
$$= (2P-1)y_i + (1-P) = E_R(z_i^2) \quad (3.3.4)$$

From Eq. (3.3.3), one finds that the transformed response $r_i = \dfrac{z_i - (1-P)}{2P-1}$ satisfies

$$E_R(r_i) = y_i,\ V_R(r_i) = \frac{P(1-P)}{(2P-1)^2} = \phi_i$$

Since the variance ϕ_i is known, one may use either $\hat{\phi}_i = \dfrac{P(1-P)}{(2P-1)^2}$ or $r_i(1-r_i)$ as an unbiased estimator of ϕ_i.

U-model:

For the U-model with known π_B, we define $x_i = \begin{cases} 1, & if\ i \in B \\ 0, & if\ i \notin B \end{cases}$ so that

$$P(z_i = 1) = P_1 y_i + (1-P_1)x_i = E_R(z_i) = E_R(z_i^2) \quad (3.3.5)$$

Here the transformed response $r_i = \dfrac{z_i - (1-P_1)x_i}{P_1}$ yields

$$E_R(r_i) = y_i \text{ and } V_R(r_i) = \frac{(1-P_1)(y_i - x_i)^2}{P_1} = \phi_i.$$ The variance ϕ_i is unknown as y_i and x_i are unknown. So, one may choose $\hat{\phi}_i = r_i(r_i - 1)$ as an estimator of ϕ_i.

F- model:

For the F-model, $E(z_i) = E(z_i^2) = y_i P + \alpha$. In this case $r_i = \dfrac{z_i - \alpha}{P}$ yields

$E_R(r_i) = y_i$ and

$$V_R(r_i) = \frac{\{P(1-P) - 2P\alpha\}y_i + \alpha(1-\alpha)}{P^2} = \phi_i \quad (3.3.6)$$

Here also ϕ_i is unknown, so we take $\hat{\phi}_i = r_i(r_i - 1)$.

3.3.2. Quantitative variables

3.3.2.1. Greenberg et al.'s (1971) model

Let y_i and x_i be the values of the sensitive and non-sensitive characteristics of y and x for the ith respondent. Denoting by z_i the RR of the ith respondent, one finds the transformed response

$r_i = \dfrac{z_i - (1-P)x_i}{P}$ yields

$$E_R(r_i) = y_i, V_R(r_i) = \frac{(1-P)(y_i - x_i)^2}{P} = \phi_i, \text{ and}$$

$$\hat{\phi}_i = (1-P)(r_i - x_i)^2. \quad (3.3.7)$$

3.3.2.2. Eriksson (1973) model

For the Eriksson model described in section 3.2, we have already shown that the transformed RR $r_i = \dfrac{z_i - (1-c)\mu_Q}{c}$ follows:

(i) $E_R(r_i) = \dfrac{E_R(z_i) - (1-c)\mu_Q}{c} = y_i$,

Table 3.1. RR models that are particular case of Arnab's (1990) model

RR models	Randomized response	r_i	ϕ_i	$\hat{\phi}_i$
		Qualitative variable: $y_i = 1(0)$ if $i \in A (i \notin A)$		
Warner (1965)	$z_i = \begin{cases} y_i & \text{with probability } P \\ 1-y_i & \text{wih probability 1 - } P \end{cases}$	$\dfrac{z_i-(1-P)}{2P-1}$	$\dfrac{P(1-P)}{(2P-1)^2}$	$\dfrac{P(1-P)}{(2P-1)^2}$
Greenberg *et al.* (1969): U- model	$z_i = \begin{cases} y_i & \text{with probability } P_1 \\ x_i & \text{with probability 1 - } P_1 \end{cases}$	$\dfrac{z_i-(1-P_1)x_i}{P_1}$	$\dfrac{(1-P_1)(y_i-x_i)^2}{P_1}$	$r_i(r_i-1)$
Mangat and Singh (1990)	$z_i = \begin{cases} y_i & \text{with probability } T \\ z_i^* & \text{with probability } 1-T \end{cases}$ where $z_i^* = \begin{cases} y_i & \text{with probability } P \\ 1-y_i & \text{with probability 1 - } P \end{cases}$	$\dfrac{z_i-(1-P^*)}{2P^*-1}$; where $P^* = T+(1-T)P$	$\dfrac{P^*(1-P^*)}{(2P^*-1)^2}$	$\dfrac{P^*(1-P^*)}{(2P^*-1)^2}$
Mangat (1994)	$z_i = y_i + (1-y_i)z_i^*$ where $z_i^* = \begin{cases} y_i & \text{with probability } P \\ 1-y_i & \text{with probability 1 - } P \end{cases}$	$\dfrac{z_i-(1-P)}{P}$	$\dfrac{(1-P)(1-y_i)}{P}$	$\dfrac{(1-P)(1-r_i)}{P}$
Brouch (1971)	$z_i = \begin{cases} y_i \text{ with probability } P \\ 1 \text{ with probability } \alpha \\ 0 \text{ with probability 1 - } P \text{ - } \alpha \end{cases}$	$\dfrac{z_i-\alpha}{P}$	$[\{P(1-P)-2\alpha P\}y_i + \alpha(1-\alpha)]/P^2$	$[\{P(1-P)-2\alpha P\}y_i + \alpha(1-\alpha)]/P^2$
		Quantitative variable		
Greenberg *et al.* (1971)	$z_i = \begin{cases} y_i & \text{with probability } P \\ x_i & \text{with probability } (1-P) \end{cases}$	$\dfrac{z_i-(1-P)x_i}{P}$	$\dfrac{(1-P)(y_i-x_i)^2}{P}$	$(1-P)(r_i-x_i)^2$
Eriksson (1973)	$z_i = \begin{cases} y_i & \text{with probability } c \\ Q_j & \text{with probability } q_j \end{cases}$	$\dfrac{z_i-(1-c)\mu_Q}{c}$	$\dfrac{(1-c)}{c}\left[(y_i-\mu_Q)^2+\dfrac{\sigma_Q^2}{c}\right]$	$(1-c)\left[(r_i-\mu_Q)^2+\dfrac{\sigma_Q^2}{c}\right]$
Additive model	$z_i = y_i + x_i$	$z_i - x_i$	σ_x^2	σ_x^2
Multiplicative model)	$z_i = y_i\vartheta$	z_i/μ_ϑ	$y_i^2\sigma_\vartheta^2/\mu_\vartheta^2$	$r_i^2\sigma_\vartheta^2/(\sigma_\vartheta^2+\mu_\vartheta^2)$
Distributive model	$z_i = x_{1i}y_i + x_{2i}$	$\dfrac{z_i-\mu_{x2}}{\mu_{x1}}$	$\dfrac{\sigma_{x1}^2y_i^2+\sigma_{x2}^2}{\mu_{x1}^2}$	$\dfrac{\sigma_{x1}^2r_i^2+\sigma_{x2}^2}{\mu_{x1}^2+\sigma_{x1}^2}$
Bar-lev *et al.* (2004)	$z_i = \begin{cases} y_i & \text{with probability } p \\ y_ix_i & \text{with probability 1-}p \end{cases}$	$\dfrac{z_i}{p+(1-p)\mu_x}$	$\dfrac{y_i^2(1-p)\begin{Bmatrix} p(1-\mu_x)^2 \\ +\sigma_x^2 \end{Bmatrix}}{\{p+(1-p)\mu_x\}^2}$	$\dfrac{(1-p)\left[p(1-\mu_x)^2+\sigma_x^2\right]r_i^2}{(1-p)\left[p(1-\mu_x)^2+\sigma_x^2\right]+\{p+(1-p)\mu_x\}^2}$
Gjestavang and Singh (2009)	$z_i = \begin{cases} y_i & \text{with probability } p_1 \\ y_ix_i & \text{with probability } p_2 \\ k & \text{with probability } p_3 \end{cases}$	$\dfrac{z_i-kp_3}{p_1+p_2\mu_x}$	$\alpha y_i^2+\beta y_i+\gamma$ where α,β and γ are given in Arnab (2025)	$\dfrac{\alpha r_i^2+\beta r_i+\gamma}{1+\alpha}$

(ii) $V_R(r_i)=\frac{(1-c)}{c}\left[(y_i-\mu_Q)^2+\frac{\sigma_Q^2}{c}\right]=\phi_i$ and

(iii) An unbiased estimator of $V_R(r_i)$ is $\hat{\phi}_i=(1-c)\left[(r_i-\mu_Q)^2+\frac{\sigma_Q^2}{c}\right]$.

A tabular representation of RR models that might be regarded as a particular case of Arnab's (1990) model was provided by Arnab (2025). It is presented here as I feel that this may be beneficial to the readers.

3.3 Estimation from the general sampling designs

Let $U=(1,..,i,..,N)$ be a finite population of N units from which a sample s of n units is selected with the probability $p(s)$ using a sampling design p. The inclusion probabilities for the i th, and i and $j(\neq i)$ th units are denoted by $\pi_i=\sum_{s\supset i}p(s)$ and $\pi_{ij}=\sum_{s\supset i,j}p(s)$ respectively. The character under study y is sensitive, and no direct response (DR) of y is available from the respondents. Our objective is to estimate the population mean $\bar{Y}=\sum_{i\in U}y_i/N$ or total $Y=\sum_{i\in U}y_i$ from the selected sample s.

3.4.1. Estimation under direct response

If direct responses y_is are available, one may define a linear homogeneous unbiased estimator for the population total as

$$\hat{Y}_{\text{lh}}=\sum_{i\in s}b_{si}y_i \tag{3.3.8}$$

where $\sum_{i\in s}$ denotes the sum over the distinct units in s and b_{si}s are unknown constants that satisfiy the unbiasedness condition

$$\sum_{s\supset i}b_{si}p(s)=1\ \forall i\in U \tag{3.3.9}$$

The variance of $\hat{Y}_{\text{lh}}$ and its unbiased estimator $\hat{\text{Var}}(\hat{Y}_{\text{lh}})$ are worked out as

$$\text{Var}(\hat{Y}_{\text{lh}})=\sum_{i\in U}(\alpha_i-1)y_i^2+\sum_{i\neq}\sum_{j\in U}(\alpha_{ij}-1)y_iy_j \tag{3.3.10}$$

and

$$\hat{\text{Var}}(\hat{Y}_{\text{lh}})=\sum_{i\in s}c_{si}y_i^2+\sum_{i\neq}\sum_{j\in s}c_{sij}y_iy_j \tag{3.3.11}$$

where

$$\alpha_i=\sum_{s\supset i}b_{si}^2p(s),\ \alpha_{ij}=\sum_{s\supset i,j}b_{si}b_{sj}p(s),$$

$$\sum_{s\supset i}c_{si}p(s)=\alpha_i-1 \text{ and } \sum_{s\supset i,j}c_{si}p(s)=\alpha_{ij}-1 \tag{3.3.12}$$

3.4.2. Estimation under RR

Since the variable of interest y is sensitive and no direct response is avail, we obtain RR responses from the respondents that satisfy the following general model proposed by Arnab (1990):

$$E_R(r_i)=y_i,\ V_R(r_i)=\phi_i \text{ and } \text{C}_R(r_i,r_j)=0 \text{ for } i\neq j \tag{3.3.13}$$

Theorem 3.4.1.

(i) $\hat{Y}_{\text{Rlh}}=\sum_{i\in s}b_{si}r_i=\hat{Y}_{\text{lh}}|_{y_i=r_i}$ is an unbiased estimator of Y

(ii) The variance of $\hat{Y}_{\text{Rlh}}$ is

$$\begin{aligned}\text{Var}(\hat{Y}_{\text{Rlh}})&=\left[\sum_{i\in U}(\alpha_i-1)y_i^2+\sum_{i\neq}\sum_{j\in U}(\alpha_{ij}-1)y_iy_j\right]+\sum_{i\in U}\alpha_i\phi_i\\&=\text{Var}(\hat{Y}_{\text{lh}})+\sum_{i\in U}\alpha_i\phi_i\end{aligned}$$

(ii) An unbiased estimator of $\text{Var}(\hat{Y}_{\text{Rlh}})$ is

$$\begin{aligned}\hat{\text{Var}}(\hat{Y}_{\text{Rlh}})&=\left[\sum_{i\in s}c_{si}r_i^2+\sum_{i\neq}\sum_{j\in s}c_{sij}r_ir_j\right]+\sum_{i\in s}b_{si}\hat{\phi}_i\\&=\hat{\text{Var}}(\hat{Y}_{\text{lh}})\Big|_{y_i=r_i}+\sum_{i\in s}b_{si}\hat{\phi}_i\end{aligned}$$

where $\hat{\phi}_i$ is an unbiased estimator of ϕ_i i.e., $E_R(\hat{\phi}_i)=\phi_i$.

Proof: Proof of the theorem is given by Arnab (2025).

3.4.3. Estimation from specific sampling strategies

3.4.1. Arbitrary sampling design

The estimator $\hat{Y}_{\text{Rlh}}$ reduces to the following Horvitz-Thompson estimator if $b_{si}=1/\pi_i$

$$\hat{Y}_{\text{HT}}(r)=\sum_{i\in s}\frac{r_i}{\pi_i} \tag{3.4.1}$$

The Theorem 3.4.1. yields

(i) $E\left[\hat{Y}_{\text{HT}}(r)\right] = Y$,

(ii) $\text{Var}\left[\hat{Y}_{\text{HT}}(r)\right] = \left[\sum_{i\in U}\frac{1-\pi_i}{\pi_i}y_i^2 + \sum_{i\neq}\sum_{j\in U}\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}y_iy_j\right] + \left(\sum_{i\in U}\frac{\phi_i}{\pi_i}\right)$ and

(iii) $\hat{\text{Var}}\left[\hat{Y}_{\text{HT}}(r)\right] = \left[\sum_{i\in s}\frac{1-\pi_i}{\pi_i^2}r_i^2 + \sum_{i\neq}\sum_{j\in s}\frac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j\pi_{ij}}r_ir_j\right] + \left(\sum_{i\in s}\frac{\hat{\phi}_i}{\pi_i}\right)$ (3.4.2)

3.4.2. Simple random sampling without replacement (SRSWOR) design

For an SRSWOR, $\pi_i = \frac{n}{N}$, $\pi_{ij} = \frac{n(n-1)}{N(N-1)}$. In this case $\hat{Y}_{\text{HT}}(r) = \frac{N}{n}\sum_{i\in s} r_i = N\overline{r}_s$ and $\overline{r}_s = \sum_{i\in s} r_i / n$. Eqs. (3.4.1) and (3.4.2) yield

(i) $E(N\overline{r}_s) = Y$,

(ii) $\text{Var}(N\overline{r}_s) = N^2\left(\frac{1}{n}-\frac{1}{N}\right)S_y^2 + \frac{N}{n}\sum_{i\in U}\phi_i$ and

(iii) $\hat{\text{Var}}(N\overline{r}_s) = N^2\left(\frac{1}{n}-\frac{1}{N}\right)s_r^2 + \frac{N}{n}\sum_{i\in s}\hat{\phi}_i$ (3.4.3)

where $\overline{r}_s = \sum_{i\in s} r_i / n$, $(N-1)S_y^2 = \sum_{i\in U}(y_i - \overline{Y})^2$ and $(n-1)s_r^2 = \sum_{i\in s}(r_i - \overline{r}_s)^2$

3.4.3. Probability proportional to size with replacement (PSWR) Sampling Scheme

For the PPSWR sampling scheme, a respondent may be selected more than once in a sample. In this case, the respondent should provide n_i RR independently. Let $r_{(j)}$ be the RR from the respondent selected at the $j(=1,..,n)$ th draw with probability $p_{(j)}$ i.e., if the jth draw selects $i(=1,..N)$th unit, then $r_{(j)} = r_i$, $p_{(j)} = p_i$ and $\phi_{(j)} = \phi_i$. The Hansen-Hurwitz estimator for the total Y is

$$\hat{Y}_{\text{HH}}(r) = \frac{1}{n}\sum_{j=1}^{n}\frac{r_{(j)}}{p_{(j)}} \tag{3.4.4}$$

For the PPSWR sampling,

(i) $E\left[\hat{Y}_{\text{HH}}(r)\right] = E_p\left[E_R\{\hat{Y}_{\text{HH}}(r)\}\right] = Y$

(ii) The variance of $\hat{Y}_{\text{HH}}(r)$ is

$$\text{Var}\left[\hat{Y}_{\text{HH}}(r)\right] = V_p\left[E_R\{\hat{Y}_{\text{HH}}(r)\}\right] + E_p\left[V_R\{\hat{Y}_{\text{HH}}(r)\}\right] = \frac{1}{n}\sum_{i\in U}p_i\left(\frac{y_i}{p_i} - Y\right)^2 + \frac{1}{n}\sum_{i\in U}\frac{\phi_i}{p_i}$$

(iii) An unbiased estimator of variance $\text{Var}\left[\hat{Y}_{\text{HH}}(r)\right]$ is

$$\hat{\text{Var}}\left[\hat{Y}_{\text{HH}}(r)\right] = \frac{1}{n(n-1)}\sum_{j=1}^{n}\left(\frac{r_{(j)}}{p_{(j)}} - \hat{Y}_{\text{HH}}(r)\right)^2$$

3.4.4. SRSWR Sampling Scheme

For the SRSWR, if $p_i = 1/N$ for all $i \in U$. Hence substituting $p_i = 1/N$ in $\hat{Y}_{\text{HH}}(r)$, $\text{Var}\left[\hat{Y}_{\text{HH}}(r)\right]$ and $\hat{\text{Var}}\left[\hat{Y}_{\text{HH}}(r)\right]$, one finds the following results:

(i) $\hat{Y}_{\text{WR}}(r) = N\,\overline{r}$ with $\overline{r} = \frac{1}{n}\sum_{j=1}^{n} r_{(j)}$ is unbiased for Y

(ii) The variance of $\hat{Y}_{\text{WR}}(r)$ is

$$\text{Var}\left[\hat{Y}_{\text{WR}}(r)\right] = N^2\frac{\sigma_y^2 + \overline{\phi}}{n}$$

where $\sigma_y^2 = \frac{1}{N}\sum_{i\in U}(y_i - \overline{Y})^2$ and $\overline{\phi} = \frac{1}{N}\sum_{i\in U}\phi_i$

and

(iii) An unbiased estimator of $\text{Var}\left[\hat{Y}_{\text{WR}}(r)\right]$ is

$$\hat{\text{Var}}\left[\hat{Y}_{\text{WR}}(r)\right] = N^2\frac{\overline{s}_r^2}{n}$$

where $\overline{s}_r^2 = \sum_{j=1}^{n}(r_{(j)} - \overline{r})^2 / (n-1)$

4. REPEATED RANDOMIZED RESPONSE

It is well known that the RR techniques provide a less efficient estimator than that of the DR technique. One way of improving the efficiency of RR techniques is to repeat the RR technique more than once with little or no additional cost. Arnab *et al.* (2016, 2017) proposed the general multiple RR method and the method of estimation of parameters from the methods. All the existing methods for multiple RR models, e.g., Franklin (1989), Kuk (1990), Odumade and Singh (2009), Singh and Grewal (2013), Chaudhuri and Christofied (2013), Singh and Sedory (2013), Hussain *et al.* (2014, 2017), Arnab and Shangodoyin (2015), Jayaraj *et al.* (2018), among others, may be considered special cases of Arnab et al.'s (2017) model.

4.1 Arnab's Multiple RR method

Arnab (2017) proposed a multiple RR model where the ith respondent produces $k(\geq 1)$ RR responses $z_{i1},..,z_{ij},...,z_{ik}$ using k RR techniques $R_1,..,R_j,...,R_k$ independently. The transformed response r_{ij} obtained from z_{ij} is assumed to follow the model

$$E_R(r_{ij}) = y_i,\ V_R(r_{ij}) = \phi_{ij} \text{ and } C_R(r_{ij}, r_{kl}) = 0 \text{ for } (i,j \neq k,l) \qquad (4.1.1)$$

From the responses z_{ij} s, a pooled response of y_i is obtained as $\overline{r}_i = \dfrac{\sum_{j=1}^{k} r_{ij}/\phi_{ij}}{\sum_{j=1}^{k} 1/\phi_{ij}}$. Then $\overline{r}_i$ follows the model:

$$E_R(\overline{r}_i) = y_i,\ V_R(\overline{r}_i) = \overline{\phi}_i \text{ and } C_R(\overline{r}_i, \overline{r}_j) = 0 \text{ for } (i \neq j) \qquad (4.1.2)$$

where $\overline{\phi}_i = \left(\sum_{j=1}^{k} 1/\phi_{ij}\right)^{-1}$.

4.1.1 Example (Arnab, 2017)

Let R_j be the W-model with the parameter P_j so that the ith respondent provides either the answer "Do you belong to the sensitive group A?" with the probability P_j or "Do you belong to the non-sensitive group $\overline{A}$?" with the probability $1-P_j$. In this case $E_R(z_{ij}) = y_i P_j + (1-y_j)(1-P_j) = E(z_{ij}^2)$, and the transform response $r_{ij} = \dfrac{z_{ij} - (1-P_j)}{2P_j - 1}$ follows the model

$E_R(r_{ij}) = y_i,\ V_R(r_{ij}) = \phi_j = P_j(1-P_j)/(2P_j-1)^2$ and $C_R(r_{ij}, r_{kl}) = 0$ for $(i,j \neq k,l)$. The average response from the ith respondent is $\overline{r}_{i\mathrm{W}} = \sum_{j=1}^{k} w_{ij} r_{ij}; w_{ij} = \dfrac{P_j(1-P_j)}{(2P_j-1)^2} / \sum_{j=1}^{k} \dfrac{P_j(1-P_j)}{(2P_j-1)^2}$. Noting that $y_i = 1$, if $i \in A$ and $y_i = 0$, if $i \notin A$, an unbiased estimator of $\pi = \sum_{i \in U} y_i / N$ under the SRSWR sampling of size n is $\hat{\pi}_\mathrm{W} = \frac{1}{n}\sum_{u=1}^{n} \overline{r}_{(u)}$ where $\overline{r}_{(u)}$ is the average transformed response from the respondent selected at the uth draw $u = 1,..,n$.

Theorem 4.1.

(i) $E(\hat{\pi}_\mathrm{W}) = \pi$

(ii) $\mathrm{Var}(\hat{\pi}_\mathrm{W}) = \dfrac{\pi(1-\pi)}{n} + \dfrac{\phi_\mathrm{W}}{n}$

(iii) An unbiased estimator of $\mathrm{Var}(\hat{\pi}_\mathrm{W})$ is

$$\mathrm{Var}(\hat{\pi}_\mathrm{W}) = \frac{1}{n(n-1)} \sum_{u=1}^{n} (\overline{r}_{(u)} - \hat{\pi}_\mathrm{W})^2$$

Proof:

Denoting $E_p(E_R)$ and $V_p(V_R)$ as expectation and variance with respect to the sampling design (RR technique), one obtains

$$\text{(i) } E(\hat{\pi}_\mathrm{W}) = E_p\left[E_R(\hat{\pi}_\mathrm{W})\right] = E_p\left(\frac{1}{n}\sum_{u=1}^{n} y_{(u)}\right) = \overline{Y} = \pi$$

and

$$\text{(ii) } \mathrm{Var}(\hat{\pi}_\mathrm{W}) = V_p\left[E_R(\hat{\pi}_\mathrm{W})\right] + E_p\left[V_R(\hat{\pi}_\mathrm{W})\right]$$
$$= V_p\left[\frac{1}{n}\sum_{u=1}^{n} y_{(u)}\right] + E_p\left[\frac{1}{n^2}\sum_{u=1}^{n} \phi_\mathrm{W}\right]$$
$$= \frac{\pi(1-\pi)}{n} + \frac{\phi_\mathrm{W}}{n}$$

(iii) Follows from the fact that $\overline{r}_{(u)}$ are iid random variables.

4.2 Extensions of Arnab's model

4.2.1. Franklin's model

In Franklins' (1989, 1989a) model, a SRSWR sample s of n respondents was selected. If the ith respondent is selected in the sample s, then the respondent is requested to draw $k(\geq 1)$ independent samples g_{ij} from the distribution g_j, $j=1,..,k$ if the respondent belongs to the sensitive group A. Otherwise, if the respondent belonged to the non-sensitive category $\overline{A}$, then the respondent should submit k independent samples h_{ij} from the distribution h_j, $j=1,..,k$. The mean and variance of $g_j(h_j)$ were known as $\mu_{gj}(\mu_{hj})$ and, $\sigma^2_{gj}(\sigma^2_{hj})$ respectively. In this case the responses can be expressed as

$$z_{ij} = y_i g_{ij} + (1-y_i)h_{ij}\ ;\ i(\in U), j=1,..,k \qquad (4.2.1)$$

For the model (4.2.1), responses z_{ij} s follow the model

$$E_R(z_{ij}) = y_i(\mu_{gj}-\mu_{hj})+\mu_{hj} \text{ and}$$
$$V_R(z_{ij}) = y_i\sigma^2_{gj}+(1-y_i)\sigma^2_{hj} \qquad (4.2.2)$$

The average of the responses (called row average) $\overline{z}_i = \frac{1}{k}\sum_{j=1}^{k} z_{ij}$ follows the model

$$E_R(\overline{z}_{i\cdot}) = (\overline{\mu}_g-\overline{\mu}_h)y_i+\overline{\mu}_h \text{ and}$$
$$V_R(\overline{z}_{i\cdot}) = \frac{1}{k}\left\{y_i\left(\overline{\sigma}^2_g-\overline{\sigma}^2_h\right)+\overline{\sigma}^2_h\right\} \qquad (4.2.3)$$

where $\overline{\mu}_g = \frac{1}{k}\sum_{j=1}^{k}\mu_{gj}, \overline{\mu}_h = \frac{1}{k}\sum_{j=1}^{k}\mu_{hj}, \overline{\sigma}^2_g = \frac{1}{k}\sum_{j=1}^{k}\overline{\sigma}^2_{gj}$

and $\overline{\sigma}^2_h = \frac{1}{k}\sum_{j=1}^{k}\overline{\sigma}^2_{hj}$

Theorem 4.2.1.

(i) $\hat{\pi}_{\text{F1}} = \frac{1}{n}\sum_{i=1}^{n}\frac{\overline{z}_{i\cdot}-\overline{\mu}_h}{\overline{\mu}_g-\overline{\mu}_h}$, $\overline{\mu}_g \neq \overline{\mu}_h$ is an unbiased estimator of π

(ii) The variance $\hat{\pi}_{\text{F1}}$ of is

$$\text{Var}(\hat{\pi}_{\text{F1}}) = \frac{\pi(1-\pi)}{n}+\frac{\pi\left(\overline{\sigma}^2_g-\overline{\sigma}^2_h\right)+\overline{\sigma}^2_h}{nk\left(\overline{\mu}_g-\overline{\mu}_h\right)^2}$$

(iii) unbiased estimator of $\text{Var}(\hat{\pi}_{\text{F1}})$ is

$$\hat{\text{Var}}(\hat{\pi}_{\text{F1}}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\overline{r}_{\text{F1}i}-\hat{\pi}_{\text{F1}})^2$$

Proof: The detailed proof is given in Arnab (2025).

4.4.2. Kuk's model

In Kuk's (1990) RR model, respondents are requested to draw $k(\geq 1)$ cards at random with replacement from a pack-1 (pack-2) of cards containing P_1 $(P_2 \neq P_1)$ proportion of black and $1-P_1$ $(1-P_2)$ proportion of red cards if he or she belonged to the sensitive group A (non-sensitive group $\overline{A}$). Let $z_{1i}(z_{2i})$ be the total number of black cards drawn by ith the respondent from pack-1 (pack-2). Then the RR from the ith respondent is given by

$$z_i = y_i z_{1i} + (1-y_i)z_{2i} \qquad (4.2.4)$$

Let $\overline{z}_i = z_i / k$ be the proportion of black cards drawn and $r_i = \frac{\overline{z}_i - P_2}{P_1 - P_2}$. Then one obtains the following RR model:

$$E_R(r_i) = y_i,\ V_R(r_i) = \frac{y_i(P_1-P_2)(1-P_1-P_2)+P_2(1-P_2)}{k(P_1-P_2)^2} = \phi_{\text{K}i}$$

and $C_R(r_i, r_j) = 0$ for $i \neq j$ (4.2.5)

It can be seen that an unbiased estimator of $\phi_{\text{K}i}$ is

$$\hat{\phi}_{\text{K}i} = \frac{r_i(P_1-P_2)(1-P_1-P_2)+P_2(1-P_2)}{k(P_1-P_2)^2} \qquad (4.2.6)$$

Eq. (4.2.5) and (4.2.6) yield

Theorem 4.2.2.

(i) $\hat{\pi}_{\text{K}} = \frac{1}{n}\sum_{i=1}^{n} r_i$ is an unbiased estimator of π

(ii) The variance of $\hat{\pi}_{\text{K}}$ is

$$\text{Var}(\hat{\pi}_{\text{K}}) = \frac{\pi(1-\pi)}{n}+\frac{\overline{\phi}_{\text{K}}}{n}$$

(iii) An unbiased estimator of $\text{Var}(\hat{\pi}_{\text{K}})$ is

$$\hat{\text{Var}}(\hat{\pi}_{\text{K}}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(r_i-\hat{\pi}_{\text{K}})^2$$

4.4.3. Singh and Grewal's model

In Singh and Grewal's (2013) model, respondents belonging to the sensitive (non-sensitive) group $A(\bar{A})$ are asked to draw a random sample from a Geometric distribution with a success probability $P_{G1}(P_{G2})$ and provide the total number of cards drawn z as his or her RR. In this case the RR from the ith respondent can be expressed as

$$z_i = z_{i1}y_i + z_{i2}(1-y_i) \tag{4.2.7}$$

and it follows:

$$E_R(z_i) = \frac{y_i}{P_{G1}} + \frac{(1-y_i)}{P_{G2}},$$

$$V_R(z_i) = \frac{1-P_{G1}}{P_{G1}^2}y_i + \frac{1-P_{G2}}{P_{G2}^2}(1-y_i) \text{ and } C_R(z_i,z_j)=0$$

for $i \neq j$ (4.2.8)

Hussain et al.'s (2014) and (2017) used the Negative Binomial distribution and the generalized geometric distribution to the Franklin (1989) model. Al-Sobhi, Hussain, and Al-Zaharai (2014) proposed two RR models using Polya's urn process, which can be regarded as a special case of Franklin's model. Dihidar (2016) considered the Hypergeometric distributions in Franklin's model. Su *et al.* (2015) considered the Franklin's model with the Bernoulli distribution.

5. RR FOR MULTIPLE ATTRIBUTES

In this section, consider a finite population U, where an individual may possess none or at least one of the $k(\geq 1)$ sensitive attributes $A_1,..,A_k$. The attributes A_js may not be disjoint; there may be some overlap among them, i.e., a person may possess more than one attribute. For example, categories are A_1: victims of domestic violence, A_2: victims of sexual abuse and A_3: Orthodox Christians. RR models for the multiple sensitive attributes having no overlap were proposed by several researchers that include Abul-Ela *et al.* (1967), Greenberg *et al.* (1989), Eriksson (1973), Drane (1976), Kim and Flueck (1978), Burke (1981, 1982, 1990), Mukhopadhyaya (1980), Mukerjee (1981), Thamane (1981), and Chang and Liang (1996), among others. RR techniques with two overlapping attributes were proposed by Christofied (2005), Moshagen and Musch (2012), Barabesi *et al.* (2012), Lee *et al.* (2013), and Hsieh *et al.* (2018). All the RR techniques mentioned above were restricted to the SRSWR method only. Arnab (2023) proved a general RR model that is applicable to any sampling design. Most of the existing models described above can be regarded as a special case of this model.

5.1 Arnab's general model

Arnab (2023) proposed a general model for the multiple attributes. In this method, the $i(\in U)$th respondent is directed to perform k randomized devices $R_1,R_2,...,R_k$ independently for each of the k attributes $A_1,A_2,...,A_k$. Let $z_i(j)$ be the response from the R_j made by the ith respondent. Here we assume that the transformed response $r_i(j)$ from $z_i(j)$ follows the model:

$$E_R(r_i(j)) = y_i(j),\ V_R(r_i(j)) = \phi_i(j) \text{ and}$$
$$Cov(r_i(j), r_k(l)) = 0 \text{ for } (i,j) \neq (k,l) \tag{5.1.1}$$

where $y_i(j)=1$ if the ith respondent possesses the atfribute A_j and $y_i(j)=0$, other wise.

Then the proportion of individuals in the population of size N that belong to the attribute A_j is

$$\pi_j = \sum_{i\in U} y_i(j)/N \tag{5.1.2}$$

Similarly, the proportion of individuals in the population that belong to the overlap

$$F = F(j_1, j_2,..., j_t) = A_{j_1} \cap A_{j_1} \cap ... \cap A_{j_1} \text{ is}$$
$$\pi_F = \sum_{i\in U} y_i(j_1)\times y_i \times (j_2)\times....\times y_i(j_t)/N;$$
$$j_1 j_2..j_t \subset 1,2,..,k$$
$$= \sum_{i\in U} F_i/N \text{ where } F_i = y_i(j_1)\times y_i(j_2)\times....\times y_i(j_t) \tag{5.1.3}$$

Denoting by

$$\hat{F}_i = r_i(j_1)\times r_i(j_2)\times....\times r_i(j_t) = \prod_{k=1}^{t} r_i(j_k), \text{ one finds}$$

$$E(\hat{F}_i) = F_i, V(\hat{F}_i) = \prod_{k=1}^{t} E(r_i^2(j_k)) - F_i^2$$

$$= \prod_{k=1}^{t} E[\phi_i(j_k) + y_i(j_k)] - F_i^2 = \phi_{F_i}$$

and $E(\hat{F}_i, \hat{F}_l) = 0$ for $i \neq l$

An unbiased estimator of $V(\hat{F}_i)$ is

$$\hat{V}(\hat{F}_i) = \hat{F}_i(1-\hat{F}_i) = \hat{\phi}_{F_i} \tag{5.1.4}$$

5.1.1. Estimation of proportion

For a general sampling design with the inclusion probabilities π_i and π_{ij} for the i th and the pair i and $j(\neq i)$ th units, the Horvitz-Thompson type estimators of π_F is

$$\hat{\pi}_{(\text{F})\text{HT}} = \frac{1}{N}\sum_{i\in s}\frac{\hat{F}_i}{\pi_i} \tag{5.1.5}$$

The properties of the estimators $\hat{\pi}_{(\text{F})\text{HT}}$ are stated as follows without derivations. Details are given in Arnab (2025).

Theorem 5.1.1.

For a general sampling design,

(i) $\hat{\pi}_{(\text{F})\text{HT}}$ is an unbiased estimator for π_F

(ii) The variance of $\hat{\pi}_{(\text{F})\text{HT}}$ is

$$\text{Var}\left[\hat{\pi}_{(\text{F})\text{HT}}\right] = \frac{1}{N^2}\left[\sum_{i\neq}\sum_{j\in U}(\pi_i\pi_j - \pi_{ij})\left(\frac{F_i}{\pi_i} - \frac{F_j}{\pi_j}\right)^2 + \sum_{i\in U}\frac{\phi_{Fi}}{\pi_i}\right];$$

(iii) An unbiased estimator of the variance of $\text{Var}\left[\hat{\pi}_{(\text{F})\text{HT}}\right]$ is

$$\hat{\text{Var}}\left[\hat{\pi}_{(\text{F})\text{HT}}\right] = \frac{1}{N^2}\left[\sum_{i\neq}\sum_{j\in s}\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\left(\frac{\hat{F}_i}{\pi_i} - \frac{\hat{F}_j}{\pi_j}\right)^2 + \sum_{i\in s}\frac{\hat{\phi}_{Fi}}{\pi_i}\right]$$

Theorem 5.1.2.

(i) For the SRSWR design, $\hat{\pi}_{(\text{F})\text{W}} = \frac{1}{n}\sum_{i=1}^{n}\hat{F}_i$ is an unbiased estimator of π_F

where $\sum_{i=1}^{n}$ denotes the sum of all the selected n units with repetition.

(ii) The variance of $\hat{\pi}_{(F)\text{W}}$ is

$$\text{Var}\left[\hat{\pi}_{(F)\text{W}}\right] = \frac{1}{n}\left[\pi_F(1-\pi_F) + \bar{\phi}_F\right];\ \bar{\phi}_F = \frac{1}{N}\sum_{i\in U}\phi_{Fi}$$

(iii) An unbiased estimator of the variance of $\hat{\pi}_{(F)\text{W}}$ is

$$\hat{\text{Var}}\left[\hat{\pi}_{(\text{F})\text{W}}\right] = \frac{1}{n(n-1)}\sum_{i=1}^{n}\left(\hat{F}_i - \hat{\pi}_{(\text{F})\text{W}}\right)^2$$

For more commentary on these subjects, readers are referred to Arnab (2025) and the detained references provided there.

6. OPTIONAL RR TECHNIQUES

The randomized response technique is used when the subject of inquiry is sensitive. But the sensitivity is highly subjective. In an optional randomized response technique (ORT), the majority of the respondents feel that the subject of the inquiry is sensitive, but a minority may think that the subject is not sensitive enough, and therefore they provide a direct response (DR) without any hesitation. An RRT that provides an option to the respondents either to provide an RR or a DR without any reservation is termed as an ORT.

ORT was introduced by Chaudhuri and Mukerjee (1988) and was extended by Gupta *et al.* (2002), Arnab (2004), Arnab and Olaomi (2019). Chaudhuri and Saha (2005), Huang (2008), Pal (2008), and Chaudhuri and Dihidar (2009), among others.

ORT is mainly classified into two categories, e.g., full and partial ORT. In the full ORT, respondents are partitioned into two groups G and $\bar{G}$. Respondents belonging to the group G felt that the subject of inquiry is sensitive, and hence, they provide only RR. While respondents belonging to the complementary group $\bar{G}$ felt that the subject of inquiry is not sensitive enough and they always provide a DR. Several researchers, including Chaudhuri and Mukerjee (1988), Arnab (2004), Arnab, *et al.*. (huss2019), and Chaudhuri and Saha (2005), considered the full ORT.

In the partial ORT, a respondent can provide either an RR or a DR depending on their assessment of the level of sensitivity at the time of providing the response. It assumed that a respondent provides an RR with an unknown probability W and a DR with a probability

$1-W$. The quantity W is termed as a measure of the sensitivity level of the inquiry, and it is estimated through the data collected. Gupta *et al.* (2002), Pal (2008), Chaudhuri and Dihidar (2009), Huang (2010), Hussain *et al.* (2014), and Kalucha *et al.* (2015), among others, proposed the partial ORT.

Most of the ORTs available in the literature are based on the SRSWR sampling. Arnab (2004, 2019) Chaudhuri *et al.* (2009) among others, provided ORT for complex survey designs and showed that the ORTs are more efficient than compulsory RRT. Arnab and Olaomi (2019) proposed a two-phase ORT that performs better than the single-phase ORT.

6.1 Estimation of the population mean

For an ORT, the ith respondent provides a randomized response z_i, if he or she belongs to the sensitive group G and a direct response y_i if he or she belongs to the non-sensitive group $\overline{G}$. Let r_i, the revised RR obtained from z_i follows the model (3.3.13). Then, the revised response for the ORT

$$\tilde{r}_i = \delta_i r_i + (1-\delta_i) y_i \text{ where } \delta_i = 1(0) \text{ if i} \in \text{G (i} \notin \text{G)} \tag{6.1.1}$$

follows the model

$$E_R(\tilde{r}_i) = y_i,\ V_R(\tilde{r}_i) = (1-\delta_i)\phi_i, \text{ and } C_R(\tilde{r}_i, \tilde{r}_j) = 0$$

for $i \neq j = 0$ (6.1.2)

The Horvitz-Thomson estimator of the population mean $\overline{Y}$ for the general sampling design with the inclusion probability π_i for the ith unit is defined as

$$\hat{\overline{Y}}_{\text{HT0}} = \frac{1}{N}\sum_{i\in s}\frac{\tilde{r}_i}{\pi_i} \tag{6.1.3}$$

The properties of the estimator are presented in the following theorem without derivation. Details are given in Arnab (2004).

Theorem 6.1.1.

For a sampling design with positive inclusion probabilities of the first and second order π_i and $\pi_{ij}\,(j \neq i)$,

(i) $\hat{\overline{Y}}_{\text{HT0}} = \frac{1}{N}\sum_{i\in s}\frac{\tilde{r}_i}{\pi_i}$ is an unbiased estimator of $\overline{Y}$

(ii) The variance of $\hat{\overline{Y}}_{\text{HT0}}$ is

$$\text{Var}(\hat{\overline{Y}}_{\text{HT0}}) = \frac{1}{N^2}\left[\sum_{i<}\sum_{j\in U}(\pi_i\pi_j - \pi_{ij})\left(\frac{y_i}{\pi_i} - \frac{y_j}{\pi_j}\right)^2 + \sum_{i\in G}\frac{\phi_i}{\pi_i}\right]$$

(iii) An unbiased estimator of the variance $\text{Var}(\hat{\overline{Y}}_{\text{HT0}})$ is

$$\hat{\text{Var}}(\hat{\overline{Y}}_{\text{HT0}}) = \frac{1}{N^2}\left[\sum_{i<}\sum_{j\in s}\frac{\pi_i\pi_j - \pi_{ij}}{\pi_{ij}}\left(\frac{\tilde{r}_i}{\pi_i} - \frac{\tilde{r}_j}{\pi_j}\right)^2 + \sum_{i\in s_G}\frac{\hat{\phi}_i}{\pi_i}\right]$$

where $\hat{\phi}_i$ is an unbiased estimator of ϕ_i and s_G is the set of respondents in the sample that belong to group G.

For an SRSWOR, $\pi_i = n/N$ and $\pi_{ij} = n(n-1)/\{N(N-1)\}$. In this case, $\overline{\tilde{r}}_{\text{o}} = \frac{1}{n}\sum_{i\in s}\tilde{r}_i$ is an unbiased estimator of $\overline{Y}$. The variance of $\overline{\tilde{r}}_{\text{o}}$ and its unbiased estimator become

$$\text{Var}(\overline{\tilde{r}}_{\text{o}}) = (1-f)\frac{S_y^2}{n} + \frac{1}{nN}\sum_{i\in G}\phi_i \text{ and}$$

$$\hat{\text{Var}}(\overline{\tilde{r}}_{\text{o}}) = (1-f)\frac{s_{\tilde{r}}^2}{n} + \frac{1}{nN}\sum_{i\in s_G}\hat{\phi}_i$$

where $S_y^2 = \frac{1}{N-1}\sum_{i\in U}(y_i - \overline{Y})^2$,

$s_{\tilde{r}}^2 = \frac{1}{n-1}\sum_{i\in s}(\tilde{r}_i - \overline{\tilde{r}})^2$ and

$$\text{Var}\left(\hat{\overline{Y}}_{\text{OT}}\right) = \frac{1}{N^2}\left[V_{\text{YG}} + E\left(V_{\text{YG}|\tilde{s}_g} + V_{\text{YG}|\tilde{s}_{\overline{g}}} + \sum_{i\in\tilde{s}_g}\frac{\phi_{Ai}}{\pi_i^2\pi_{gi|\tilde{s}}}\right)\right].$$

7. NONRANDOMIZED RESPONSE TECHNIQUE

The RR techniques are used to collect highly sensitive data that are not obtainable through the DR. Hence, RR techniques reduce non-responses and provide good quality data. However, a few researchers, e.g., Geurts (1980), Dalton & Metzger (1992), Hubbard, Casper & Lessler (1989), Tian *et al.* (2009), Rueda *et al.* (2018), and Arnab *et al.* (2019), among others, identified certain drawbacks: the interviewers need to explain the techniques clearly to the respondents; the respondents very often misunderstand the logic behind the RR techniques. As a result, the RR technique becomes more expensive and time-consuming than the DR and not applicable to the mail inquiry. Respondents often provide different responses if they repeat the experiment again. Furthermore, the RR method possesses lower efficiency than the DR. To overcome such difficulties, the nonrandomized response (NRR) technique has been proposed by Swensson (1974), Takahasi and Sakasegawa (1977), Tian *et al.* (2007, 2009), and Yu *et al.* (2008), among others. In the NRR methods, randomization devices are replaced with non-sensitive or innocuous questions. Hence, the proposed methods bear the reproducibility property, are cheaper to implement, are easily understandable to the respondents, are feasible for face-to-face inquiry, and, above all, maintain respondents' privacy. Some of the NRR techniques are described below without derivation of the results. Details are obtainable from Arnab (2025).

7.1 Takahasi and Sakasegawa method:

The Takahasi and Sakasegawa (1977) method is also known as the implicit RR technique. In this method, three independent samples s_1, s_2 and s_3 of sizes n_1, n_2, and n_3 are selected by the SRSWR method. The population is also divided into three mutually exclusive and exhaustive non-sensitive (or innocuous) categories: B_1: born in the 1st quarter (January-April); B_2: born in the 2nd quarter (May-August), and B_3: born in the third quarter (September -December) of the year. Respondents are requested to answer "yes" or "no" to the membership of the combinations of the sensitive $A(\bar{A})$ and non-sensitive attributes according to the following plan.

Categories	Sample s_1		sample s_2		Sample s_3	
	Attribute		Attribute		Attribute	
	A	$\bar{A}$	A	$\bar{A}$	A	$\bar{A}$
B_1	no	yes	yes	no	yes	no
B_2	yes	no	no	yes	yes	no
B_3	yes	no	yes	no	no	yes

Details of the method of estimation of π_A the proportion of persons belonging to the sensitive category A are given in Chaudhuri and Mukerjee (1988), and Arnab (2025).

7.2 Triangular model

The triangular model was introduced by Tian *et al.* (2007, 2009), and Yu *et al.* (2008).

Let, $y_i = \begin{cases} 1, \text{ if the ith unit} \in \text{A} \\ 0, \text{ if the ith unit} \in \bar{\text{A}} \end{cases}$ and

$$x_i = \begin{cases} 1, \text{ if the ith unit} \in B \\ 0, \text{ if the ith unit} \in \bar{B} \end{cases}$$

where A is sensitive; $\bar{A}$, complementary to A and non-sensitive; B is non-sensitive; $\bar{B}$ is complementary to B and non-sensitive; A and B are independent. A sample of n respondents is selected by the SRSWR method. Each of the respondents was requested to put a tick mark against their membership to one of the two groups, $\bar{A} \cap \bar{B}$ $(y=0, x=0)$ or its complementary group $(A \cap \bar{B}) \cup A$ $((y=0, x=1) \cup (y=1))$, in a circle. Otherwise, the respondent should put a solid dot in a triangle. Group $\bar{A} \cap \bar{B}$ is non-sensitive, while $(A \cap \bar{B}) \cup A$ is a mixture of sensitive and non-sensitive individuals. So, the privacy of the respondent is preserved, as the respondents would not identify whether the respondent belongs to the sensitive group A or not.

7.2.1. Estimation of proportion π_y

Let π_y and π_x be the proportions of individuals belonging to the groups A and B, respectively. Out of the selected n respondents, let n_1 of them put a tick mark in the circle, and the remaining $n - n_1$ put a solid dot in a triangle. Noting that n_1 follows a binomial distribution with parameters n and $\theta_t = (1-\pi_y)(1-\pi_x)$, we obtain the following theorem:

Theorem 7.2.1.

(i) $\hat{\pi}_{\mathrm{T}y} = 1 - \dfrac{\hat{\theta}_t}{1-\pi_x}$ is an unbiased estimator of π_y

where $\hat{\theta}_t = n_1 / n$

(ii) The variance of $\hat{\pi}_{\mathrm{T}y}$ is

$$\mathrm{Var}(\hat{\pi}_{\mathrm{T}y}) = \frac{\pi_y\left(1-\pi_y\right)}{n} + \frac{\pi_x\left(1-\pi_y\right)}{n\left(1-\pi_x\right)}$$

(iii) An unbiased estimator of $\mathrm{Var}(\hat{\pi}_{\mathrm{T}y})$

$$\hat{\mathrm{Var}}(\hat{\pi}_{\mathrm{T}y}) = \frac{1}{(n-1)}\frac{\hat{\theta}_t(1-\hat{\theta}_t)}{(1-\pi_x)^2}$$

7.3 Crosswise model

Tian *et al.* (2007, 2009) and Yu *et al.* (2008) also proposed the crosswise method. In this method, the respondents are asked put a tick mark (√) if he or she belongs to one of the groups $(y=0,x=0)\cup(y=1,x=1)$ and $(y=0,x=1)\cup(y=1,x=0)$. Otherwise, the respondent should put a cross mark ($\times$) as his or her response. Here both the groups are a mixture of the sensitive and non-sensitive attributes A and $\overline{A}$. Hence, the respondent's privacy is maintained.

Let n_1 be the total number of √ marks obtained from a sample of n respondents elected by the SRSWR method. Then, n_1 follows a binomial distribution with parameters n and $\theta_c = (1-\pi_y)(1-\pi_x)+\pi_x\pi_y$. In this case, one obtains the following results, analogous to theorem 7.2.1.

Theorem 7.3.1.

(i) $\hat{\pi}_{\mathrm{C}y} = \dfrac{\hat{\theta}_c-(1-\pi_x)}{2\pi_x-1}$ is an unbiased estimator of π_y where $\hat{\theta}_c = n_1/n$

(ii) The valiance of $\hat{\pi}_{cy}$ is

$$\mathrm{Var}(\hat{\pi}_{cy}) = \frac{\pi_y(1-\pi_y)}{n} + \frac{\pi_x(1-\pi_x)}{n(2\pi_x-1)^2}$$

(iii) An unbiased estimator of $\mathrm{Var}(\hat{\pi}_{\mathrm{C}y})$

$$\hat{\mathrm{Var}}(\hat{\pi}_{\mathrm{C}y}) = \frac{1}{(n-1)}\frac{\hat{\theta}_c(1-\hat{\theta}_c)}{(1-\pi_x)^2}$$

7.4 Parallel model

Tian (2014) proposed the parallel model, which is also known as the NRR version of the unrelated question model to estimate π_A, the proportion of the persons belonging to a certain sensitive group A. Group $\overline{A}$, the complement to group A, may or may not be sensitive. Two non-sensitive attributes, B and C, with the known prevalences π_B and π_C, are chosen. The sensitive attribute A and the non-sensitive attributes B and C are assumed to be mutually independent.

A sample of n individuals was selected by the SRSWR method, and each of them was asked to answer "yes" or "no" to one of the two questions, Q1 or Q2. The respondent should answer Q1 if he or she possesses the attribute $\overline{C}$ (complement of C), and answer Q2 if he or she possesses the attribute C. Questions Q1 and Q2 are as follows:

Q1: Do you possess the attribute B?

Q2: Do you possess the attribute A?

The privacy of the respondent is protected, as the respondent did not reveal the question he or she answered. The probability of obtaining a "yes" answer from a respondent is

$$\theta_P = \left(1-\pi_C\right)\pi_B + \pi_C\pi_A$$

Let $\hat{\theta}_P$ be the proportion of "yes" answers in the sample, then we have the following results:

Theorem 7.4.1.

(i) $\hat{\pi}_P = \dfrac{\hat{\theta}_P-\left(1-\pi_C\right)\pi_B}{\pi_C}$ is an unbiased estimator of π_A.

(ii) The variance of $\hat{\pi}_A$ is

$$\mathrm{Var}\left(\hat{\pi}_P\right) = \frac{\theta_{\mathrm{P}}(1-\theta_{\mathrm{P}})}{n\pi_C^2}$$

(iii) An unbiased estimator of $\mathrm{Var}\left(\hat{\pi}_P\right)$ is

$$\hat{\mathrm{Var}}\left(\hat{\pi}_P\right) = \frac{\hat{\theta}_P\left(1-\hat{\theta}_P\right)}{(n-1)\pi_C^2}$$

8. MEASURES OF PRIVACY PROTECTION

Answering sensitive questions in the direct interviewing method may result in jeopardizing respondents' privacy. The main purpose of the RR technique is to mask the true response of a respondent deliberately so that respondent's true response remains undisclosed to the interviewer or third parties. While the statistician's interest is to collect reliable data to obtain efficient estimates of the survey parameters of interest. The two objectives, maintaining adequate privacy of the respondents and improving the efficiency of the survey estimates, are conflicting with each other. Several measures of privacy protections (PP) are proposed in the literature, but they are far from adequate as they exclude several important factors like the degree of complexity of the RR devices, total cost, level of embarrassment, and cooperation from the respondents (see Arnab, 2025). Most of the measures are applicable to the SRSWR sampling only. Important measures of privacy protection (PP) are due to Anderson (1975), Leysieffer and Warner (1976), Lanke (1976), Flinger, Policello, and Singh (1976), and Quatember (2009), among others. Bhargava and Singh (1999, 2000, 2002) and Guerriero and Sandri (2007) compared performances of several RR techniques. They concluded that Warner's (1965) model and Mangat and Singh's (1990) are equally efficient. Arnab (2021, 2023) argued that the number of RR trials for all the RR models are not the same, e.g., Warner's (1965) model requires only one trial (drawing only one card), and Kuk's model requires drawing $c(\geq 1)$ cards. Hence, the Warner's model is comparable with Kuk's model if and only if $c=1$. The number of trials for Mangat and Singh (1994), and Singh and Grewal's (2013) RR techniques are not fixed; they are random variables. Hence, one should compare different RR models incorporating the level of PP and expected number of trials fixed to a certain level. A good review is given by Chaudhuri and Mukerjee (1988), Bhargava and Singh (1999, 2000, 2002), Guerriero and Sandri (2007), Quatember (2012), and Arnab (2017, 2025).

ACKNOWLEDGEMENTS

The author wishes to thank the referee for fruitful suggestions.

REFERENCES

Abernathy, J.R., Greenberg, B.G., and Horvitz, D.G. (1970). Estimates of induced abortion in urban North Carolina. *Demography*, 7, 19-29.

Abul-Ela, A.L.A., Greenberg, B.D. and Horvitz, D.G. (1967). Multi-proportions randomized response model. *Journal of the American Statistical Association*, 62, 990-1008.

Anderson, H. (1975). Efficiency versus protection in RR designs for estimating proportions, *Technical Report* 9, University of Lund, Sweden.

Arnab, R. (1990). On commutativity of design and model expectations in randomized response surveys. *Communications in Statistics e Theory and Methods*, 19, 3751-3757.

Arnab, R. (1996). Randomized response trials: A unified approach for qualitative data. *Communications in Statistics - Theory and Methods*, 25(6), 1173-1183.

Arnab, R. (2004). Optional randomized response techniques for complex survey designs. *Biometrical Journal*, 46, 1, 114-124

Arnab, R. (2017). *Survey Sampling Theory and Applications*, Academic Press, U.K.

Arnab, R. (2021). Randomized Response Trial using Geometric Distribution. *Journal of the Indian Society of Agricultural Statistics,*75, 195–204.

Arnab, R. (2023). Dalenius and Vitale randomized response technique for complex survey designs. *Communications in Statistics-Theory and Methods*, 52, 684-692.

Arnab, R. (2025). *Indirect methods of data collection and analysis.* Singapore, Springer.

Arnab, R. and Olaomi, J.O. (2019). Optional randomized response technique for quantitative characteristics. *Communications in Statistics-Theory and Methods*, 48, 4154-4170.

Arnab, R. and Shangodoyin, D.K. (2015). Randomized response technique using maximum likelihood estimator. *Communications in Statistics-Theory and Methods*, 44, 3340-3349.

Arnab, R., and Rueda, M. (2016). Optional randomized response: A critical review. In vol. 34 of *Hand Book of Statistics*, ed. A. Chaudhuri, T.C. Christofieds and C.R.Rao, 253-71. Oxford, U.K., Elsevier.

Arnab, R., and Shangodoyin, D.K., and Arcos, A. (2019). Nonrandomized response model for complex survey designs. *Statistics in Transitions*, 20, 67-

Arnab, R., Shangodoyin, D.K. and Dwivedi, V.K. (2016). Generalization of Randomized Response Techniques for Complex Survey Designs. *Journal of the Indian Society of Agricultural Statistics,*70, 131-138.

Arnab, R., Shangodoyin, D.K. and Kgosi, P.M. (2017). Randomized Response Techniques with Multiple Responses. *Journal of the Indian Society of Agricultural Statistics*, 71, 201-205.

Barabesi, L., Franceschi, S., Marcheselli, M. (2012). A randomized response procedure for multiple sensitive attributes. *Statistical Papers*, 53, 703-718.

Bhargava, M. and Singh, R. (1999). On relative efficiency of certain randomized response strategies. *Journal of the Indian Society of the Agricultural Statistics*, 52, 245-253.

Bhargava, M. and Singh, R. (2000). A modified randomization device for Warner's model. *Statistica Anno* LX(2), 315-321.

Bhargava, M. and Singh, R. (2002). On the efficiency comparison of certain randomized response strategies. *Metrika*, 55, 191-197.

Boruach, R.F. (1972). Relations among statistical methods for assuring confidentiality of social research. *Communications in Statistics e Theory and Methods*, 19, 3751-3757.

Bourch, R. (1971). Assuring Confidentiality of Responses in Social Research: A Note on Strategies. *The American Sociologist,* 6, 308-411.

Burke, P.D. (1981). On the analysis of some multivariate randomized response designs for categorical data. *Journal of Statistical Planning and Inference*, **5**, 165-170.

Burke, P.D. (1982). Randomized response multivariate designs for categorical data. *Communications in Statistics-Theory and Methods,* **11**, 2889-2901.

Burke, P.D. (1990). Estimating a distribution for each category of sensitive variate. *Communications in Statistics-Theory and Methods,* **19**, 3233-3241.

Chang, H.-J. and Liang, D.-H. (1996). A two-stage unrelated randomized response procedure. *Australian Journal of Statistics*, **38**, 43-58.

Chang, H.-J., Wang, C.-L., and Huang, K.-C. (2004). On estimating the proportion of a qualitative sensitive character using randomized response sampling. *Quality and Quantity*, **38**, 675-680.

Chaudhuri, A. and Dihidar, K. (2009). Estimating means of stigmatizing qualitative and quantitative variables from discretionary responses randomized or direct. *Sankhyā*, Volume **71-B**, 123-136

Chaudhuri, A. and Mukerjee, R. (1988). *Randomized response: Theory and Techniques*. Marcel Dekker, New York.

Chaudhuri, A. and Saha, A. (2005). Optional versus compulsory randomized response techniques in complex surveys. *Journal of the statistical planning and inference*, **135**, 516-527.

Chaudhuri, A., and Christofides , T.C. (2013). *Indirect Questioning in Sample surveys.* Berlin, Springer

Christofied, T.C. (2015). A new version of the item count technique. *Model Assisted Statistics and Applications*, **10**, 289-297.

Cobo, B., Rueda, M., and Torrecillas, F. (2016). Application of randomized response techniques for investigating cannabis use by Spanish university students. *International Journal of Methods in Psychiatric Research*, DOI: 10.1002/mpr.1517

Dalton, D. R. & Metzger, M. (1992). Integrity testing for personal selections: An unsparing perspective. *Journal of Business Ethics*, **12**, 147-156.

Dihidar, K. (2016). Estimating sensitive population proportion by generating randomized response following direct and inverse hypergeometric distribution. Hand book of Statistics -34, Chaudhuri, A ., Christofides, T.C. and Rao, C.R., 443-469.

Drane, W. (1976). On theory of randomized responses of two sensitive questions. *Communications in Statistics-Theory and Methods,* **5**, *565-575.*

Duffy, C.J., and Waterton, J.J. (1984). RR models for estimating the distribution function of a quantitative character. *International Statistical Review*, **52**, 165-171

Eichorn, B.H., and Hayre, L.S. (1983). Scrambled randomized response method for obtaining sensitive quantitative data. *Journal of Statistical Planning and Inference,* **7**, 307-316.

Eriksson, S.A. (1973). A new model for randomized response. *International Statistical Review*, **41**, 101-113

Flinger, M.A., Policello, G.E. and Singh, J. (1977). A comparison of two RR survey methods with consideration for the level of respondent protection, *Communications in Statistics-Theory and Methods.* **6**, 1511-1524.

Folsom, R.E., Greenberg, B.G., Horvitz, D.G., and Abernathy. J.R. (1973). The two alternative questions randomized response model for human surveys. *Journal of the American Statistical Association*, **68**, 525-530.

Fox, J. A. and Tracy, P.E. (1986). *Randomized response: A method for sensitive surveys.* London, Sage.

Franklin, L.A. (1989). A comparison of estimators for randomized response sampling with continuous distribution from a dichotomous population. *Communications in Statistics-Theory and Methods*, **18**, 489-505.

Franklin, L.A. (1989a). A comparison of estimators for randomized response sampling with continuous distribution from dichotomous populations. *Communications in Statistics-Theory and Methods*, **18**, 489-505.

Geurts, M.D. (1980). Using a randomized response research design to determine non-response. *Journal of the Academy of Marketing Science*, **8**, 83-90.

Gjestvang, C.R. and Singh, S. (2009). An improved randomized response model: estimation of mean. *Journal of Applied Statistics*, **36**, 1361-1367.

Goodstadt, M. and Gruson, V. (1975). The Randomized Response Technique: A Test of Drug Use. *Journal of the American Statistical Association,* **70**, 814-818

Greenberg, B. G., Kuebler, R.R., Abernathy J.R. and Horvitz, D.G. (1971). Application of the randomized 21 response technique in obtaining quantitative data. *Journal of the American Statistical Association.* **66**, 243-250.

Greenberg, B.D., Abul-Ela, A.L.A., Simmons, W.R., Horvitz, D.G. (1989). The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association*, **64**, 520-539

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

Gupta, S., Gupta, B., Singh, S. (2002). Estimation of sensitivity level of personal interview survey question. *Journal of Statistical Planning and Inference* , **100**, 239-247

Horvitz, D. G., Greenberg, B.G. and Aber Abernathy, J.R. (1976). RR: a data gathering device for sensitive questions. *International Statistical Review*, **44**, 181-196.

Horvitz, D. G., Shah, B. V., & Simmons, W. R. (1967). The U- question randomized response model. *In Proceedings of the survey research methods section of the American Statistical Association* (pp. 65- 72).

Hsieh, S-H., Lee, S-M. and Tu, S-H. (2018). Randomized response technique for a multilevel attribute using a single sensitive question. *Statistical Papers,* **59**, *291-306.*

Huang, K.C., (2008). Estimation of sensitive characteristics using optional randomized technique, *Qual. Quant.*, **42**, 679 - 686

Huang, K.C., (2010). Unbiased estimators of mean, variance and sensitivity level for quantitative characteristics in finite population sampling. *Metrika,* **71**, 341-352.

Hubbard, M.L., Casper, R. A., and Lessler, J.T. (1989). Respondents OCO reactions to item count lists and randomized response. *In Proceedings of Survey Research Section, American Statistical Association*, Washington D.C. 544OCo448.

Hussain, Z., Shabbir, J., Al-zahrani, B., and Khan, M. (2014). On using negative binomial distribution as a randomization device in sensitive surveys. *Communications in Statistics-Simulation and Computation*, **45**, 3584-3596.

Hussain, Z., Shabbir, J., Pervez, Z. Shah, S.F. and Khan, M. (2017). Generalized geometric distribution of order k: A flexible choice to randomized response. *Communications in Statistics-Simulation and Computation,* **46**, 4708-4721.

Jayraj, A. Sedory, S.A., Singh, S. and Odumade, O. (2018). A new estimator of proportion with a linear function using data from two-decks randomized response model. *Communications in Statistics- Theory and Methods*, **38**, 439-446

Kalucha, G., Gupta, S., Dass, B. K. (2015). Ratio estimation of finite population mean using optional randomized response models. *Journal of Statistical Theory and practice*, **9**, 633–645.

Kim, J.I. and Flueck, J.A. (1978). An additive randomized response model. In: *Proceedings of Social Statistics section, American Statistical Association,* 351-355.

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

Lanke, J. (1975). On the choice of unrelated questions in Simmons' version of RR. *Journal of the American Statistical Association*, **68**, 525-530

Lara, D., Strickler, J., D´iaz olavarrieta, C., Ellertson, C. (2004). Measuring induced abortion in mexico a comparison of four methodologies, *Sociological methods & research*, **32**(4), 529-558 DOI 10.1177/0049124103262685

Lee, C., Sedory, S.A., Singh, S. (2013). Estimating at least seven measures of qualitative variables from a single sample using randomized response technique. *Statistical probability letters*, 83, 399-409.

Leysieffer, R.W., and Warne, S.L. (1976). Respondent jeopardy and optimal designs in RR models *Journal of the American Statistical Association*, **71**, 649-656.

Mangat, N. S., and Singh, R. (1990). An alternative randomized response procedure. *Biometrika*, **77**, 349- 442.

Mangat, N.S. (1994). An improved Randomized response strategy. *Journal of the Royal Statistical Society, Series B*, 56, 93-95.

Mangat, N.S., Singh, R. and Singh, S. (1992). An improved U- question randomized response strategies. *Calcutta Statistical Association Bulletin*, **42**, 277-281.

Mielecka-Kubień, Z. and Toniszewski, M. Estimation of illicit drug use among high school students in the Silesian voivodship (Poland) with the use of the randomized response technique. Mathematical population studies, **29**(2), 47–57, https://doi.org/10.1080/08898480.2021.1893009

Mieth, L., Mayer, M., Hoffmann A., Buchner, A., and Bell, R. (2021). Do they really wash their hands? Prevalence estimates for personal hygiene behaviour during the COVID-19 pandemic based on indirect questions. Mieth *et al.* BMC Public Health, **21**(12). https://doi.org/10.1186/s12889-020- 10109-5

Moshagen, M. and Musch, J. (2012). Surveying multiple sensitive attributes using extension of randomized response technique. *International Journal of Public Opinion research*, **24**, 508-523.

Odumade, O., and Singh, S. (2009). Efficient use of two deck of cards in randomized response sampling. *Communications in Statistics - Theory and Methods*, **38**, 439-446

Pal, S. (2008). Unbiasedly estimating the total of a stigmatizing variables from a complex survey on permitting options for direct or randomized responses. *Statistical papers*, **49**, 157-164.

Pollock, K.H., and Bek, Y. (1976). A comparison of three RR models for quantitative data. *Journal of the American Statistical Association*, **71**, 884-886.

Poole, W.K. (1974). Estimation of the distribution function for a continuous type random variable through RR. *Journal of the American Statistical Association*, **69**,1002-1005.

Quatember, A. (2009). A standardization of randomized response strategies. *Survey Methodology*, **35**,143- 152.

Quatember, A. (2012). An extension of the standardized randomization technique to a multi-stage setup. *Statistical Methods and Applications*, **21**, 475-484.

Rueda, M., Perri, P.F., and Cobo, B. (2018). Advances in estimating by the item sum technique using auxiliary information in complex surveys. AStA Advances in Statistical Analysis **102**, 455-478.

Singh, J. (1976). A note on RR techniques. *In Proceedings of the survey research methods section of the American Statistical Association* (p. 772).

Singh, J. (1978). A note on maximum likelihood estimation from randomized response models. *In Proceedings of the survey research methods section of the American Statistical Association* (pp. 282- 283).

Singh, S. (1993). An alternative to Warner's randomized response technique. *Statistica*, *anno* LIII, 67-71.

Singh, S. and Joarder, A.H. (1997). Unknown repeated trials in randomized response sampling. *Journal of the Indian Society of Agricultural Statistics*, **50**, 70-74.

Singh, S. and Sedory, S.A. (2013). A new randomization device for sensitive characteristics: an application to negative hypergeometric distribution. *Metron*, **71**, 3-8.

Singh, S. and Singh, R. (1992). Improved Franklin's model for randomized response sampling. *Journal of the Indian Society of Agricultural Statistics*, **30**, 109-122.

Singh, S. and Singh, R. (1993). Generalized Franklin's model for randomized response sampling. *Communications in Statistics -Theory and Methods*, **22**, 741-755.

Singh, S., and Grewal, I. (2013). Geometric distribution as randomization device: Implemented to the Kuk's model. *International journal of contemporary Mathematical Sciences*, **8**, 243-248.

Singh, S., Horn, H., Singh, R. and Mangat, N.S. (2003). On the use of modified randomization device for estimating the prevalence of the sensitive attribute. *Statistics in Transition*, **6**, 515-522.

Srivastava, R., Nigam, A.K., and Singh, N. (2015). Application of Randomized Response Techniques in Estimation of prevalence of Child Sexual Abuse. *Statistics and Applications* {ISSN 2454-7395(online)}

Su, S., Sedory, S.A. and Singh, S. (2015). Kuk's model adjusted for protection and efficiency. *Sociological Methods and Research*, **44**, 534-551.

Swensson, B. (1974). Combined questions: a new survey technique for eliminating evasive answer bias (I). basic theory. *Report No. 70 of errors in survey research project, Institute of Statistics, University of Stockholm.*

Takahasi, K. and Sakasegawa, H. (1977). A randomized response technique without making use of randomizing device. *Annals of the Institute of Statistical Mathematics*, **29**,1-8.

Tian, G.-L. (2014). A new non-randomized response model: The parallel model. *Statistica Neerlandica,* **68**, 294- 323.

Tian, G.-L., Tan, M., and Ng, K.W. (2007). An exact non-iterative sampling procedure for discrete missing data problems. *Statistica Neerlandica*, **61**, 232-242.

Tian, G.-L., Yu, J.W., Tang, M.L., Geng, Z. (2007). A new nonrandomized model for analyzing sensitive questions with binary outcomes. *Statistics in Medicine*, **26**, 4238-4252.

Tian, M.T., Tian, G.L., Tang, M.L. (2009). Sample surveys with sensitive questions: a non-randomized response approach. *American Statistician*, **63**, 9-16.

Warner, S.L. (1971). The linear randomized response model. *Journal of the American Statistical Association*, **366**, 884-888.

Yu, J.-W., Tian, G.-L., Tang, M.-L. (2008). Two new models for survey sampling with sensitive characteristics: design and Analysis. *Metrika*, **67**, 251-263.