STATISTICS IN TRANSITION new series, December 2017
Vol. 18, No. 4, pp. 669–686, DOI 10.21307/stattrans-2017-006

A GENERALIZED RANDOMIZED RESPONSE MODEL

Housila P. Singh, Swarangi M. Gorey[1]

ABSTRACT

In this paper we have suggested a generalized version of the Gjestvang and Singh (2006) model and have studied its properties. We have shown that the randomized response models due to Warner (1965), Mangat and Singh (1990), Mangat (1994) and Gjestvang and Singh (2006) are members of the proposed RR model. The conditions are obtained in which the suggested RR model is more efficient than the Warner (1965) model, Mangat and Singh (1990) model and Mangat (1994) model and Gjestvang and Singh (2006) model. A numerical illustration is given in support of the present study.

Key words: sensitive variable, population proportion, Gjestvang and Singh's model, variance, efficiency.
AMS Subject Classification: 62D05.

1. Introduction

The collection of data through personal interviews surveys on sensitive issues such as induced abortions, alcohol and drug abuse (Weissman et al., 1986, Fisher et al., 1992) as well as on attitudes (Antonak and Livnech, 1995), on sexual behaviour (Williams and Suen, 1994, Jarman, 1997) and family income is a serious issue. Warner (1965) introduced an ingenious technique known as the randomized response technique for estimating the proportion π of people bearing a sensitive attribute, say A, in a given community from which a sample is collected. For estimating π, a simple random sample of n respondents is selected with replacement from the population. For collecting information on the sensitive characteristic, Warner (1965) made use of randomization device. The randomization device consists of a deck of cards with each card having one of the following two statements:

(i) I belong to sensitive group A;
(ii) I do not belong to sensitive group A,

[1] School of Studies in Statistics, Vikram University, Ujjain-456010, M.P., India.

represented with probabilities p_0 and $(1-p_0)$ respectively in the deck of cards. Each respondent in the sample is asked to select a card at random from the well-shuffled deck. Without showing the card to the interviewer, the interviewee answers the question, "Is the statement true for you?" the number of respondents n_1 that answer "yes" is binomially distributed with parameters $p_0\pi+(1-p_0)(1-\pi)$. The maximum likelihood estimator π exists for $p_0 \neq \frac{1}{2}$ and is given by

$$\hat{\pi}_w = \frac{(n_1/n)-(1-p_0)}{(2p_0-1)} \tag{1.1}$$

which is unbiased and has the variance

$$V(\hat{\pi}_w) = \frac{\pi(1-\pi)}{n} + \frac{p_0(1-p_0)}{n(2p_0-1)^2}. \tag{1.2}$$

Mangat and Singh (1990) envisaged a two-stage randomized response model. In the first stage, each respondent was requested to use a randomization device, R_1, such as a deck of cards with each card containing one of the following two statements: (i) "I belong to sensitive group A", (ii) "Go the randomization device R_2". The statements occur with probabilities T_0 and $(1-T_0)$, respectively, in the first device R_1. In the second stage, if directed by the outcome of R_1, the respondent is requested to use the randomization device R_2, which is the same as the Warner (1965) device. Under the two-stage randomized response model, an unbiased estimator of the population proportion π, due to Mangat and Singh (1990) is given by

$$\hat{\pi}_{ms} = \frac{(n_1/n)-(1-T_0)(1-p_0)}{(2p_0-1)+2T_0(1-p_0)} \tag{1.3}$$

with the variance

$$V(\hat{\pi}_{ms}) = \frac{\pi(1-\pi)}{n} + \frac{(1-p_0)(1-T_0)[1-(1-p_0)(1-T_0)]}{n[2p_0-1+2T_0(1-p_0)]^2} \tag{1.4}$$

Mangat (1994) investigated another randomized response model where each respondent selected in the sample was requested to report "yes" if he/she belonged to the sensitive group A; otherwise, he/she was instructed to use the

Warner (1965) device. Under this model, Mangat (1994) obtained an unbiased estimator of the population proportion π given by

$$\hat{\pi}_m = \frac{(n_1/n) - (1 - p_0)}{p_0} \tag{1.5}$$

with the variance

$$V(\hat{\pi}_m) = \frac{\theta_m(1 - \theta_m)}{np_0^2}, \tag{1.6}$$

where $\theta_m = \pi + (1 - \pi)(1 - p_0)$.

It is to be mentioned that the Mangat (1994) RR model is more efficient than both the Warner (1965) and Mangat and Singh (1990) models.

A rich growth of literature on randomized response procedure has been accumulated in Chaudhuri and Mukherjee (1987, 1988). Further, a detailed review on randomized response sampling can be found in Singh (2003). Some related work on the randomized response sampling can be also be found in Odumade and Singh (2008, 2009a, 2009b, 2010) Bouza et al. (2010) and Chaudhuri et al. (2016).

It is noted that the Mangat (1994) model has been improved by Gjestvang and Singh (2006). In this paper we have made an effort to suggest a generalized randomized response model which includes Warner (1965), Mangat and Singh (1990), Mangat (1994), Gjestvang and Singh (2006) randomized response model. It has been shown that the proposed model is superior to the models suggested by Warner (1965), Mangat and Singh (1990), Mangat (1994) and Gjestvang and Singh (2006) under some realistic conditions. Numerical illustration is given in support of the present study.

2. Suggested Randomized Response Model

In this section we propose a generalized randomized response model. For estimating π, the proportion of respondents in the population belonging to the sensitive group A, a simple random sample of n respondents is selected with replacement from the population. If the person who is selected in the sample belongs to the sensitive group A, then he or she is requested to use the randomization device R_1 that is described below. Similar to Gjestvang and Singh (2006), let α_1 and β_1 be any two positive real numbers such that $p = \alpha_1/(\alpha_1 + \beta_1)$ is the probability in the randomization device R_1 directing the selected respondent to report a scrambled response (or indirect response) as

$(1+w_1\beta_1 S_1)$, and $(1-p)=\beta_1/(\alpha_1+\beta_1)$ is the probability in the randomization device R_1 directing the selected respondent to report a scrambled response as $(1-w_1\alpha_1 S_1)$, where w_1 is a known real number and S_1 is any non-directional scrambling variable, i.e. S_1 can take positive, zero and negative values. If the person who is selected in the sample does not belong to the sensitive group A, then he or she is requested to use the randomization device R_2 that is described below. Let α_2 and β_2 be any two positive real numbers (similar to Gjastvang and Singh (2006)) such that $T=\alpha_2/(\alpha_2+\beta_2)$ is the probability in the randomization device R_2 directing the selected respondent to report a scrambled response $w_2\beta_2 S_2$, and let $(1-T)=\beta_2/(\alpha_2+\beta_2)$ be the probability in the randomization device R_2 directing the selected respondent to report scrambled response as $-w_2\alpha_2 S_2$, where w_2 is a known real number and S_2 is any non-directional scrambling variables. The main difference from the existing randomization response models is that here the distribution of the scrambling variables S_1 and S_2 may or may not be known. Gjestvang and Singh (2006) have noted that the negative response will not disclose the privacy of any respondent belonging to non-sensitive or sensitive group because they come from both groups. Here we also note that if the mean θ_i and variance γ_i^2 of the ith scrambling variable S_i (i=1,2) are known before start of the survey, then in such a situation, the value of w_i may be the function of the known quantities (θ_i, γ_i^2), i=1,2.

Theorem 2.1 An unbiased estimator of the population proportion π is given by

$$\hat{\pi}_{HS}=\frac{1}{n}\sum_{i=1}^{n} y_i \tag{2.1}$$

Proof The observed response in the proposed method has the distribution

$$y_i=\begin{cases} 1+w_1\beta_1 S_1 \text{ with probability } p\pi, \\ 1-w_1\alpha_1 S_1 \text{ with probability } (1-p)\pi, \\ w_2\beta_2 S_2 \quad \text{ with probability } T(1-\pi), \\ -w_2\alpha_2 S_2 \text{ with probability } (1-T)(1-\pi). \end{cases} \tag{2.2}$$

Let E_1 and E_2 denote the expected values over all possible samples and over the randomization device. Then we have

$$E(\hat{\pi}_{HS})=E_1E_2(\hat{\pi}_{HS})$$

$$= \frac{1}{n} E_1 \sum_{i=1}^{n} E_2(y_i). \tag{2.3}$$

where

$$y_i = \pi p(1 + w_1\beta_1 S_1) + (1-p)\pi(1 - w_1\alpha_1 S_1) + T(1-\pi)\beta_2 S_2 w_2 - w_2\alpha_2(1-T)(1-\pi)S_2$$

Let $E_2(S_1) = \theta_1$ and $E_2(S_2) = \theta_2$. Then we have

$$E_2(y_i) = \pi p(1 + w_1\beta_1\theta_1) + (1-p)\pi(1 - w_1\alpha_1\theta_1) + T(1-\pi)\beta_2\theta_2 w_2 - w_2\alpha_2(1-T)(1-\pi)\theta_2,$$

$$= \pi\{p(1 + w_1\beta_1\theta_1) + (1-p)(1 - w_1\alpha_1\theta_1)\} + (1-\pi)w_2\theta_2(T\beta_2 - (1-T)\alpha_2),$$

$$= \pi\left\{1 + w_1\theta_1\left[\frac{\alpha_1\beta_1}{(\alpha_1+\beta_1)} - \frac{\alpha_1\beta_1}{(\alpha_1+\beta_1)}\right]\right\} + (1-\pi)w_2\theta_2\left\{\frac{\alpha_2\beta_2}{(\alpha_2+\beta_2)} - \frac{\alpha_2\beta_2}{(\alpha_2+\beta_2)}\right\},$$

$$= \pi + (1-\pi),$$

$$= \pi. \tag{2.4}$$

Putting (2.4) in (2.3) we get

$$E(\hat{\pi}_{HS}) = \frac{1}{n} E_1 \sum_{i=1}^{n} \pi$$
$$= \pi$$

which proves the theorem.

Theorem 2.2 The variance of the estimator $\hat{\pi}_{HS}$ is given by

$$V(\hat{\pi}_{HS}) \frac{\pi(1-\pi)}{n} + \frac{1}{n}\left\{\pi w_1^2\alpha_1\beta_1\left(\gamma_1^2 + \theta_1^2\right) + (1-\pi)w_2^2\alpha_2\beta_2\left(\gamma_2^2 + \theta_2^2\right)\right\} \tag{2.5}$$

Proof The responses are independent, thus the variance of the estimator $\hat{\pi}_{HS}$ is given by

$$V(\hat{\pi}_{HS}) = V\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)$$
$$= \frac{1}{n^2}\sum_{i=1}^{n} V(y_i). \tag{2.6}$$

Let V_1 and V_2 denote the variance over all possible samples and the variance over the randomization device respectively. Then we have

$$V(y_i) = E_1 V_2(y_i) + V_1 E_2(y_i)$$
$$= V_1 V_2(y_i) + V_1(\pi)$$

$$= E_1 V_2(y_i). \tag{2.7}$$

Let the variance of the scrambling variables be $V(S_1) = \gamma_1^2$ and $V(S_2) = \gamma_2^2$. Then

$$V_2(y_i) = E_2(y_i^2) - \{E_2(y_i)\}^2$$
$$= \pi\{pE_2(1 + w_1\beta_1 S_1)^2 + (1-p)E_2(1 - w_1\alpha_1 S_1)^2\}$$
$$+ (1-\pi)\{TE_2(w_2\beta_2 S_2)^2 + (1-T)E_2(-w_2\alpha_2 S_2)^2\} - \pi^2,$$

$$= \pi\{p\left[1 + w_1^2\beta_1^2(\gamma_1^2 + \theta_1^2) + 2w_1\beta_1\theta_1\right]$$
$$+ (1-p)\left[1 + w_1^2\alpha_1^2(\gamma_1^2 + \theta_1^2) - 2w_1\alpha_1\theta_1\right]\}$$
$$+ (1-\pi)\{T\, w_2^2\beta_2^2(\gamma_2^2 + \theta_2^2) + (1-T)w_2^2\alpha_2^2(\gamma_2^2 + \theta_2^2)\} - \pi^2,$$

$$= \pi(1-\pi) + \pi\left[\frac{w_1^2\alpha_1\beta_1^2(\gamma_1^2 + \theta_1^2)}{(\alpha_1 + \beta_1)} + \frac{2w_1\alpha_1\beta_1\theta_1}{(\alpha_1 + \beta_1)} + \frac{w_1^2\alpha_1^2\beta_1(\gamma_1^2 + \theta_1^2)}{(\alpha_1 + \beta_1)} - \frac{2w_1\alpha_1\beta_1\theta_1}{(\alpha_1 + \beta_1)}\right]$$
$$+ (1-\pi)\left\{\frac{w_2^2\alpha_2\beta_2^2(\gamma_2^2 + \theta_2^2)}{(\alpha_2 + \beta_2)} + \frac{\beta_2 w_2^2\alpha_2^2(\gamma_2^2 + \theta_2^2)}{(\alpha_2 + \beta_2)}\right\} - \pi^2,$$

$$= \pi(1-\pi) + \frac{\pi w_1^2\alpha_1\beta_1(\gamma_1^2 + \theta_1^2)(\alpha_1 + \beta_1)}{(\alpha_1 + \beta_1)} + \frac{(1-\pi)w_2^2(\gamma_2^2 + \theta_2^2)\alpha_2\beta_2(\alpha_2 + \beta_2)}{(\alpha_2 + \beta_2)}$$

$$= \pi(1-\pi) + \pi w_1^2\alpha_1\beta_1(\gamma_1^2 + \theta_1^2) + (1-\pi)w_2^2(\gamma_2^2 + \theta_2^2)\alpha_2\beta_2. \tag{2.8}$$

Thus from (2.6), (2.7) and (2.8) we have

$$V(\hat{\pi}_{HS}) = \frac{\pi(1-\pi)}{n} + \frac{1}{n}\{\pi w_1^2\alpha_1\beta_1(\gamma_1^2 + \theta_1^2) + (1-\pi)w_2^2(\gamma_2^2 + \theta_2^2)\alpha_2\beta_2\} \tag{2.9}$$

which proves the theorem.

Corollary 2.1 Assuming that

$$\frac{\alpha_2\beta_2}{\alpha_1\beta_1} = \frac{\gamma_1^2 + \theta_1^2}{\gamma_2^2 + \theta_2^2},$$

$(\gamma_2^2 + \theta_2^2) = 1$ [similar to Gjestvang and Singh (2006,p.525)] and $w_1 = w_2 = w$ (say), the variance of the estimator $\hat{\pi}_{HS}$ in (2.5) reduces to

$$V(\hat{\pi}_{HS}) = \frac{\pi(1-\pi)}{n} + \frac{w^2\alpha_2\beta_2}{n}. \quad (2.10)$$

Proof is simple so omitted.

The variance in (2.10) of the proposed estimator $\hat{\pi}_{HS}$ can be estimated as

$$\hat{V}(\hat{\pi}_{HS}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(y_i - \hat{\pi}_{HS})^2. \quad (2.11)$$

It should be remembered here that (w, α_2, β_2) are known quantities in the variance expression (2.10). As mentioned in Gjestvang and Singh (2006), we also show that models due to Warner (1965), Mangat and Singh (1990), Mangat (1994) and Gjestvang and Singh (2006) are special cases of the suggested RR procedure (model). If we set

(i) $\{p(1+w_1\beta_1\theta_1)+(1-p)(1-w_1\alpha_1\theta_1)\} = p_0$ and

$w_2\{T\beta_2\theta_2 - (1-T)\alpha_2\theta_2\} = (1-p_0)$,

(ii) $\{p(1+w_1\beta_1\theta_1)+(1-p)(1-w_1\alpha_1\theta_1)\} = (1-p_0)(1-T_0)$ and

$w_2\{T\beta_2\theta_2 - (1-T)\alpha_2\theta_2\} = \{1-(1-p_0)(1-T_0)\}$,

(iii) $\{p(1+w_1\beta_1\theta_1)+(1-p)(1-w_1\alpha_1\theta_1)\} = 1$ and

$w_2\{T\beta_2\theta_2 - (1-T)\alpha_2\theta_2\} = (1-p_0)$

(iv) $(w_1, w_2) = (1,1)$

the proposed RR model respectively reduces to the Warner (1965), Mangat and Singh (1990), Mangat (1994) and Gjestvang and Singh (2006) models.

3. Efficiency Comparison

In the proposed procedure, if we set $w_1 = w_2 = 1$, then the procedure investigated by Gjestvang and Singh (2006, sec.2, p.524) becomes special case (or member of the present proposed procedure).

In the Gjestvang and Singh (2006) model, the observed response has the distribution

$$y_i^* = \begin{cases} 1+\beta_1 S_1 & \text{with probability } p\pi, \\ 1-\alpha_1 S_1 & \text{with probability } (1-p)\pi, \\ \beta_2 S_2 & \text{with probability } T(1-\pi), \\ -\alpha_2 S_2 & \text{with probability } (1-T)(1-\pi). \end{cases} \quad (3.1)$$

This can be also obtained just by putting $w_1 = w_2 = 1$ in (2.2).

An unbiased estimator of π due to Gjestvang and Singh (2006) is given by

$$\hat{\pi}_{GS} = \frac{1}{n}\sum_{i=1}^{n} y_i^* . \tag{3.2}$$

The variance of $\hat{\pi}_{GS}$ is given by

$$V(\hat{\pi}_{GS}) = \frac{\pi(1-\pi)}{n} + \frac{1}{n}\left\{\pi\left(\gamma_1^2+\theta_1^2\right)\alpha_1\beta_1 + (1-\pi)\left(\gamma_2^2+\theta_2^2\right)\alpha_2\beta_2\right\}. \tag{3.3}$$

Assuming that

$$\frac{\alpha_2\beta_2}{\alpha_1\beta_1} = \frac{\gamma_1^2+\theta_1^2}{\gamma_2^2+\theta_2^2}$$

and $\gamma_2^2+\theta_2^2 = 1$, then variance of $\hat{\pi}_{GS}$ in (3.3) reduces to

$$V(\hat{\pi}_{GS}) = \frac{\pi(1-\pi)}{n} + \frac{\alpha_2\beta_2}{n} . \tag{3.4}$$

From (2.5) and (3.3) we have

$$V(\hat{\pi}_{GS}) - V(\hat{\pi}_{HS}) = \frac{1}{n}\left[\pi\left(\gamma_1^2+\theta_1^2\right)\alpha_1\beta_1\left(1-w_1^2\right) + (1-\pi)\left(\gamma_2^2+\theta_2^2\right)\alpha_2\beta_2\left(1-w_2^2\right)\right] \tag{3.5}$$

which is positive if

$$\left(1-w_i^2\right) > 0 \text{, i=1,2;}$$

i.e. if

$$-1 < w_i < 1 \text{, i=1,2}$$

$$\text{i.e. if } |w_i| < 1 \text{, i=1,2}$$

Thus, we established the following theorem.

Theorem 3.1 The proposed estimator $\hat{\pi}_{HS}$ (i.e. proposed procedure) is always better than Gjeatvang and Singh's (2006) estimator $\hat{\pi}_{GS}$ (i.e. Gjestvang and Singh's (2006) procedure) if

$$|w_i| < 1 \text{, i=1,2.} \tag{3.6}$$

Further, from (2.10) and (3.4) we have

$$V(\hat{\pi}_{GS}) - V(\hat{\pi}_{HS}) = \frac{\alpha_2\beta_2}{n}\left(1-w^2\right) \tag{3.7}$$

which is always positive if

$$1-w^2>0$$

$$\text{i.e. if} \quad |w|<1. \tag{3.8}$$

Thus, we established the following corollary.

Corollary 3.1 Under the assumption

$$\frac{\alpha_2\beta_2}{\alpha_1\beta_1}=\frac{\gamma_1^2+\theta_1^2}{\gamma_2^2+\theta_2^2}$$

and

$$w_1=w_2=w \text{ (a real number, say).}$$

The proposed estimator $\hat{\pi}_{HS}$ is more efficient than Gjestvang and Singh's (2006) estimator $\hat{\pi}_{GS}$ if $|w|<1$.

Assume that the values of $(\alpha_i,\beta_i,\theta_i,\gamma_i^2, i=1,2)$ are predetermined before conducting the survey and are assumed to be known. Note that θ_1 and θ_2 are non-directional. From (1.2) and (2.10) we have that $V(\hat{\pi}_{HS})<V(\hat{\pi}_w)$ if

$$w\alpha_2\beta_2<\frac{p_0(1-p_0)}{(2p_0-1)^2} \tag{3.9}$$

which is free from the parameter π under investigation and depends on the parameters of the randomization devices. We also note that the condition (3.9) is also very flexible.

From (1.4) and (2.10) we have

$$V(\hat{\pi}_{ms})-V(\hat{\pi}_{HS})=\frac{1}{n}\left[\frac{(1-p_0)(1-T_0)[1-(1-p_0)(1-T_0)]}{\{2p_0-1+2T_0(1-p_0)\}^2}-w\alpha_2\beta_2\right]$$

which is positive if

$$w\alpha_2\beta_2<\frac{(1-p_0)(1-T_0)[1-(1-p_0)(1-T_0)]}{[2p_0-1+2T_0(1-p_0)]^2}. \tag{3.10}$$

This condition is also free from the parameter π under investigation and depends on the parameters of the randomization devices.

Further, from (1.6) and (2.10) we have that $V(\hat{\pi}_m)<V(\hat{\pi}_{HS})$ if

$$w\alpha_2\beta_2<\frac{(1-p_0)(1-\pi)}{p_0} \tag{3.11}$$

Thus, the proposed RR model is more efficient than Warner's (1965) model, Mangat and Singh's (1990) model and Mangat's (1994) model as long as the conditions (3.9), (3.10) and (3.11) are respectively satisfied.

4. Some Members of the Proposed Procedure

I. Assume that the values of $\alpha_1, \beta_1, \alpha_2, \beta_2, \theta_1, \gamma_1^2, \theta_2$ and γ_2^2 are predetermined before conducting the survey and are assumed to be known. Note that θ_1 and θ_2 are non-directional [see Gjestvang and Singh (2006), sec.3, p.525)]. In our model, if we take $w_1 = \left(\frac{2\gamma_1\theta_1}{\gamma_1^2+\theta_1^2}\right)^{1/2}$ and $w_2 = \left(\frac{2\gamma_2\theta_2}{\gamma_2^2+\theta_2^2}\right)^{1/2}$ in (2.2), then the observed response has the distribution:

$$y_{i(1)} = \begin{cases} 1+\left(\dfrac{2\gamma_1\theta_1}{\gamma_1^2+\theta_1^2}\right)^{1/2}\beta_1 S_1 \text{ with probability } p\pi, \\ 1-\left(\dfrac{2\gamma_1\theta_1}{\gamma_1^2+\theta_1^2}\right)^{1/2}\alpha_1 S_1 \text{ with probability } (1-p)\pi, \\ \left(\dfrac{2\gamma_2\theta_2}{\gamma_2^2+\theta_2^2}\right)^{1/2}\beta_2 S_2 \quad \text{ with probability } T(1-\pi), \\ -\left(\dfrac{2\gamma_2\theta_2}{\gamma_2^2+\theta_2^2}\right)^{1/2}\alpha_2 S_2 \text{ with probability } (1-T)(1-\pi). \end{cases} \tag{4.1}$$

Thus, an unbiased estimator of the population proportion π is given by

$$\hat{\pi}_{HS(1)} = \frac{1}{n}\sum_{i=1}^{n} y_{i(1)} . \tag{4.2}$$

Putting $w_1 = \left(\frac{2\gamma_1\theta_1}{\gamma_1^2+\theta_1^2}\right)^{1/2}$ and $w_2 = \left(\frac{2\gamma_2\theta_2}{\gamma_2^2+\theta_2^2}\right)^{1/2}$ in (2.5) we get the variance of $\hat{\pi}_{HS(1)}$ as

$$V(\hat{\pi}_{HS(1)}) = \frac{\pi(1-\pi)}{n} + \frac{1}{n}\left[2\gamma_1\theta_1\alpha_1\beta_1\pi + 2\gamma_2\theta_2\alpha_2\beta_2(1-\pi)\right]. \tag{4.3}$$

From (3.3) and (4.3) we have

$$V(\hat{\pi}_{GS}) - V(\hat{\pi}_{HS(1)}) = \frac{1}{n}\left[\pi\alpha_1\beta_1(\gamma_1 - \theta_1)^2 + (1-\pi)\alpha_2\beta_2(\gamma_2 - \theta_2)^2\right] \quad (4.4)$$

which is always positive provided $\gamma_1 \neq \theta_2$ and $\gamma_2 \neq \theta_2$. Thus, the proposed RR model (4.1) is always better than the RR model (3.1) due to Gjestvang and Singh (2006). In the situation where $\gamma_i = \theta_i$,(i=1,2), both the models are equally efficient.

II. If $w_1 = \frac{\theta_1}{\sqrt{\theta_1^2 + \gamma_1^2}}$ and $w_2 = \frac{\theta_2}{\sqrt{\theta_2^2 + \gamma_2^2}}$ in (2.2), then the observed response has the distribution:

$$y_{i(2)} = \begin{cases} 1 + \frac{\theta_1}{\sqrt{\theta_1^2 + \gamma_1^2}}\beta_1 S_1 \text{ with probability } p\pi, \\ 1 - \frac{\theta_1}{\sqrt{\theta_1^2 + \gamma_1^2}}\alpha_1 S_1 \text{ with probability } (1-p)\pi, \\ \frac{\theta_2}{\sqrt{\theta_2^2 + \gamma_2^2}}\beta_2 S_2 \quad \text{ with probability } T(1-\pi), \\ -\frac{\theta_2}{\sqrt{\theta_2^2 + \gamma_2^2}}\alpha_2 S_2 \quad \text{ with probability } (1-T)(1-\pi). \end{cases} \quad (4.5)$$

Thus, an estimator of the population proportion π is given by

$$\hat{\pi}_{HS(2)} = \frac{1}{n}\sum_{i=1}^{n} y_{i(2)} . \quad (4.6)$$

Inserting $w_1 = \frac{\theta_1}{\sqrt{\theta_1^2 + \gamma_1^2}}$ and $w_2 = \frac{\theta_2}{\sqrt{\theta_2^2 + \gamma_2^2}}$ in (2.5) we get the variance of (4.6) as

$$V(\hat{\pi}_{HS(2)}) = \frac{\pi(1-\pi)}{n} + \frac{1}{n}\left\{\pi\alpha_1\beta_1\theta_1^2 + (1-\pi)\alpha_2\beta_2\theta_2^2\right\} \quad (4.7)$$

From (3.3) and (4.7) we have

$$V(\hat{\pi}_{GS}) - V(\hat{\pi}_{HS(2)}) = \frac{1}{n}\left[\pi\alpha_1\beta_1\gamma_1^2 + (1-\pi)\alpha_2\beta_2\gamma_2^2\right] \quad (4.8)$$

which is always positive. Thus, the RR model proposed in (4.5) is superior to Gjestvang and Singh's (2006) RR model (3.1).

Assuming that

$$\frac{\alpha_2\beta_2}{\alpha_1\beta_1}=\frac{\theta_1^2}{\theta_2^2}, \tag{4.9}$$

the variance of $\hat{\pi}_{HS(2)}$ reduces to

$$V\left(\hat{\pi}_{HS(2)}\right)=\frac{\pi(1-\pi)}{n}+\frac{\alpha_2\beta_2\theta_2^2}{n} \tag{4.10}$$

From (3.4) and (4.10) we have

$$\begin{aligned} V\left(\hat{\pi}_{GS}\right)-V\left(\hat{\pi}_{HS(2)}\right) &= \frac{\alpha_2\beta_2}{n}\left(1-\theta_2^2\right) \\ &> 0 \text{ if } \theta_2^2<1. \end{aligned} \tag{4.11}$$

Thus, the proposed estimator $\hat{\pi}_{HS(2)}$ is more efficient than the Gjestvang and Singh (2006) estimator $\hat{\pi}_{GS}$ as long as the condition $\theta_2^2<1$ satisfied.

III. If we set $w_1=\frac{\gamma_1}{\sqrt{\theta_1^2+\gamma_1^2}}$ and $w_2=\frac{\gamma_2}{\sqrt{\theta_2^2+\gamma_2^2}}$ in (2.3), then the observed response has the distribution:

$$y_{i(3)}\begin{cases} 1+\dfrac{\gamma_1}{\sqrt{\left(\theta_1^2+\gamma_1^2\right)}}\beta_1 S_1 & \text{with probability } p\pi, \\ 1-\dfrac{\gamma_1}{\sqrt{\left(\theta_1^2+\gamma_1^2\right)}}\alpha_1 S_1 & \text{with probability } (1-p)\pi, \\ \dfrac{\gamma_2}{\sqrt{\theta_2^2+\gamma_2^2}}\beta_2 S_2 & \text{with probability } T(1-\pi), \\ \dfrac{-\gamma_2}{\sqrt{\theta_2^2+\gamma_2^2}}\alpha_2 S_2 & \text{with probability } (1-T)(1-\pi). \end{cases} \tag{4.12}$$

Thus, an unbiased estimator of the population proportion π is defined by

$$\hat{\pi}_{HS(3)}=\frac{1}{n}\sum_{i=1}^{n}y_{i(3)}. \tag{4.13}$$

Putting $w_1 = \frac{\gamma_1}{\sqrt{\left(\theta_1^2 + \gamma_1^2\right)}}$ and $w_2 = \frac{\gamma_2}{\sqrt{\left(\theta_2^2 + \gamma_2^2\right)}}$ in (2.5) we get the variance of the estimator $\hat{\pi}_{HS(3)}$ as

$$V\left(\hat{\pi}_{HS(3)}\right) = \frac{\pi(1-\pi)}{n} + \frac{1}{n}\left\{\pi\alpha_1\beta_1\gamma_1^2 + (1-\pi)\alpha_2\beta_2\gamma_2^2\right\}. \tag{4.14}$$

From (3.3) and (4.14) we have

$$V\left(\hat{\pi}_{GS}\right) - V\left(\hat{\pi}_{HS(3)}\right) = \frac{1}{n}\left\{\pi\alpha_1\beta_1\theta_1^2 + (1-\pi)\alpha_2\beta_2\theta_2^2\right\} \tag{4.15}$$

which is always positive. Thus, it follows from (4.15) that the proposed estimator $\hat{\pi}_{HS(3)}$ is more efficient than Gjestvang and Singh's (2006) estimator $\hat{\pi}_{GS}$, i.e. the RR model suggested in (4.9) is superior to the RR model in (3.1) due to Gjestvang and Singh (2006).

Assuming that

$$\frac{\alpha_2\beta_2}{\alpha_1\beta_1} = \frac{\gamma_1^2}{\gamma_2^2}, \tag{4.16}$$

the variance of the estimator $\hat{\pi}_{HS(3)}$ in (4.14) is reduced to

$$V\left(\hat{\pi}_{HS(3)}\right) = \frac{\pi(1-\pi)}{n} + \frac{\alpha_2\beta_2\gamma_2^2}{n}. \tag{4.17}$$

It can be seen from (3.4) and (4.17) that

$$V\left(\hat{\pi}_{GS}\right) - V\left(\hat{\pi}_{HS(3)}\right) = \frac{\alpha_2\beta_2}{n}\left(1-\gamma_2^2\right)$$

which is positive if

$$\gamma_2^2 < 1 \tag{4.18}$$

Thus, the proposed estimator $\hat{\pi}_{HS(3)}$ is more efficient than Gjestvang and Singh's (2006) estimator $\hat{\pi}_{GS}$ as long as the condition (4.18) is satisfied.

Remark 4.1 For $w_i = \frac{\theta_i^2}{\left(\theta_i^2 + \gamma_i^2\right)}$, (i=1,2) $w_i = \frac{\gamma_i^2}{\left(\theta_i^2 + \gamma_i^2\right)}$, (i=1,2) and $w_i = \frac{\left|\left(\theta_i^2 - \gamma_1^2\right)\right|}{\left(\theta_i^2 + \gamma_i^2\right)}$ one can get the randomized response models always better than Gjestvan and Singh's (2006) randomized response models.

Many more suitable choices of w_1 and w_2 can be considered (which may be either the function of $(\theta_i, \gamma_i, i = 1,2)$ or not) for which we can obtain the model superior to the Gjestvang and Singh (2006).

5. Relative Efficiency

It is assumed that the values of $\alpha_1, \beta_1, \alpha_2, \beta_2, \theta_1, \gamma_1^2, \theta_2$ and γ_2^2 are known before the start of the survey. It is to be noted that the Mangat (1994) model remains more efficient than the Mangat and Singh (1990) model. Also, Gjestvang and Singh (2006) have proved that the estimator $\hat{\pi}_{GS}$ proposed by them can always be made more efficient than the Warner (1965), Mangat and Singh (1990) and Mangat (1994) estimators for various choices of known parameters of the model. Thus, it is acceptable to compare the proposed model only with Gjestvang and Singh (2006).

To see the magnitude of the gain efficiency of the suggested randomized response model, we compute the percent relative efficiency (PRE) of the proposed estimator $\hat{\pi}_{HS}$ with respect to Gjestvang and Singh's (2006) estimator $\hat{\pi}_{GS}$ as follows.

$$PRE(\hat{\pi}_{HS}, \hat{\pi}_{GS}) = \frac{V(\hat{\pi}_{GS})}{V(\hat{\pi}_{HS})} \times 100 \tag{5.1}$$

or equivalently (by using (2.9) and (3.3) in (5.1))

$$PRE(\hat{\pi}_{HS}, \hat{\pi}_{GS}) = \frac{\left[\pi(1-\pi) + \left\{\pi\alpha_1\beta_1\left(\gamma_1^2 + \theta_1^2\right) + (1-\pi)\alpha_2\beta_2\left(\gamma_2^2 + \theta_2^2\right)\right\}\right]}{\left[\pi(1-\pi) + \left\{\pi w_1^2\alpha_1\beta_1\left(\gamma_1^2 + \theta_1^2\right) + (1-\pi)w_2^2\alpha_2\beta_2\left(\gamma_2^2 + \theta_2^2\right)\right\}\right]} \times 100 \tag{5.2}$$

Further, for the simplicity we have assumed $\left(\gamma_1^2 + \theta_1^2\right) = \left(\gamma_2^2 + \theta_2^2\right) = 1$ [similar to Gjestvang and Singh (2006), p.526] and $w_1 = w_2 = w$ (a real constant) under these assumptions, the $PRE(\hat{\pi}_{HS}, \hat{\pi}_{GS})$ in (5.2) reduces to :

$$PRE(\hat{\pi}_{HS}, \hat{\pi}_{GS}) = \frac{\left[\pi(1-\pi) + \left\{\pi\alpha_1\beta_1 + (1-\pi)\alpha_2\beta_2\right\}\right]}{\left[\pi(1-\pi) + w^2\left\{\pi\alpha_1\beta_1 + (1-\pi)\alpha_2\beta_2\right\}\right]} \times 100 \tag{5.3}$$

We have computed the $PRE(\hat{\pi}_{HS}, \hat{\pi}_{GS})$ by using (5.3) for $\pi = 0.05, 0.1(0.1)0.9$, and for three sets of α_i's, β_i's (i=1,2) values as (i) $\alpha_1 = 0.6, \beta_1 = 0.4, \alpha_2 = 0.3, \beta_2 = 0.7$ (ii) $\alpha_1 = 0.8, \beta_1 = 0.2, \alpha_2 = 0.4,$

$\beta_2 = 0.6$ (iii) $\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 0.5$ $|w| = 0.05, 0.1(0.1)0.9$. Findings are compiled in Table 5.1.

Table 5.1. The percent relative efficiency of the proposed model with respect to Gjestvang and Singh's (2006) model

$\alpha_1 = 0.6, \beta_1 = 0.4, \alpha_2 = 0.3, \beta_2 = 0.7$

π \ $\lvert w \rvert$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.05	104.85	104.81	104.65	104.40	104.05	103.60	103.06	102.42	101.70	100.89
0.1	104.81	108.51	108.23	107.77	107.13	106.32	105.34	104.21	102.94	101.53
0.3	104.65	119.93	119.21	118.03	116.41	114.40	112.03	109.36	106.43	103.29
0.5	104.40	133.35	132.02	129.85	126.94	123.38	119.29	114.79	110.00	105.04
0.7	104.05	160.55	157.66	153.06	147.06	140.00	132.24	124.11	115.89	107.80
0.9	293.17	288.97	273.32	250.69	224.65	198.18	173.24	150.80	131.20	114.35

$\alpha_1 = 0.8, \beta_1 = 0.2, \alpha_2 = 0.4, \beta_2 = 0.6$

π \ $\lvert w \rvert$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.05	102.90	102.87	102.78	102.64	102.43	102.16	101.84	101.46	101.03	100.54
0.1	105.22	105.17	105.01	104.74	104.36	103.87	103.28	102.60	101.82	100.95
0.3	112.66	112.56	112.13	111.42	110.45	109.23	107.77	106.10	104.23	102.19
0.5	121.56	121.36	120.58	119.30	117.55	115.38	112.84	109.97	106.84	103.50
0.7	139.58	139.16	137.53	134.89	131.37	127.10	122.24	116.96	111.40	105.71
0.9	225.60	223.49	215.43	203.21	188.27	172.00	155.57	139.79	125.15	111.86

$\alpha_1 = \beta_1 = \alpha_2 = \beta_2 = 0.5$

π \ $\lvert w \rvert$	0.05	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.05	104.37	104.34	104.20	103.98	103.66	103.25	102.76	102.19	101.54	100.80
0.1	107.92	107.85	107.59	107.17	106.58	105.84	104.94	103.90	102.72	101.42
0.3	119.42	119.25	118.56	117.42	115.87	113.93	111.65	109.07	106.24	103.20
0.5	133.22	132.89	131.58	129.45	126.58	123.08	119.05	114.61	109.89	104.99
0.7	161.16	160.42	157.53	152.95	146.96	139.92	132.18	124.07	115.87	107.79
0.9	294.21	289.98	274.19	251.39	225.17	198.54	173.47	150.94	131.27	114.38

It is observed from Table 5.1 that the values of $\text{PRE}(\hat{\pi}_{HS}, \hat{\pi}_{GS})$ are larger than 100 for the given parametric values. It follows that the suggested estimator $\hat{\pi}_{HS}$ can always be made more efficient than Gjestvang and Singh's (2006) estimator $\hat{\pi}_{GS}$ and hence more efficient than the Warner (1965), Mangat and

Singh (1990) and Mangat (1990) estimators. For larger values (or even moderately large values) of $|w|$ and π, the considerable gain in efficiency is observed by using the proposed estimator $\hat{\pi}_{HS}$ over Gjestvang and Singh's (2006) estimator $\hat{\pi}_{GS}$. Thus, we see that the proposed procedure is an improvement over Gjestvang and Singh's (2006) procedure.

We have further computed the percent relative efficiencies (PRE's) of the proposed estimators $\hat{\pi}_{HS(1)}, \hat{\pi}_{HS(2)}$ and $\hat{\pi}_{HS(3)}$ with respect to Gjestvang and Singh's (2006) estimator $\hat{\pi}_{GS}$ by using the formulae:

$$PRE\left(\hat{\pi}_{HS(1)}, \hat{\pi}_{GS}\right) = \frac{\left[\pi(1-\pi) + \pi\left(\gamma_1^2 + \theta_1^2\right)\alpha_1\beta_1 + (1-\pi)\left(\gamma_2^2 + \theta_2^2\right)\alpha_2\beta_2\right]}{\left[\pi(1-\pi) + 2\left\{\gamma_1\theta_1\alpha_1\beta_1\pi + \gamma_2\theta_2\alpha_2\beta_2(1-\pi)\right\}\right]} \times 100 \tag{5.4}$$

$$PRE\left(\hat{\pi}_{HS(2)}, \hat{\pi}_{GS}\right) = \frac{\left[\pi(1-\pi) + \pi\left(\gamma_1^2 + \theta_1^2\right)\alpha_1\beta_1 + (1-\pi)\left(\gamma_2^2 + \theta_2^2\right)\alpha_2\beta_2\right]}{\left[\pi(1-\pi) + \pi\alpha_1\beta_1\theta_1^2 + (1-\pi)\alpha_2\beta_2\theta_2^2\right]} \times 100 \tag{5.5}$$

$$PRE\left(\hat{\pi}_{HS(2)}, \hat{\pi}_{GS}\right) = \frac{\left[\pi(1-\pi) + \pi\left(\gamma_1^2 + \theta_1^2\right)\alpha_1\beta_1 + (1-\pi)\left(\gamma_2^2 + \theta_2^2\right)\alpha_2\beta_2\right]}{\left[\pi(1-\pi) + \pi\alpha_1\beta_1\gamma_1^2 + (1-\pi)\alpha_2\beta_2\gamma_2^2\right]} \times 100 \tag{5.6}$$

for $\left(\theta_1, \gamma_1^2\right) = (0.6, 0.50)$, $\left(\theta_2, \gamma_2^2\right) = (0.8, 0.36)$, $(\alpha_1, \beta_1) = (0.6, 0.4)$, $(\alpha_2, \beta_2) = (0.05, 0.95)$ [similar to Gjestvang and Singh (2006), section 4, p.527]. Findings are given in Table 5.2.

Table 5.2. The percent relative efficiencies of $\hat{\pi}_{HS(1)}, \hat{\pi}_{HS(2)}$ and $\hat{\pi}_{HS(3)}$ with respect to Gjestvang and Singh's (2006) estimator $\hat{\pi}_{GS}$

π	$PRE\left(\hat{\pi}_{HS(1)}, \hat{\pi}_{GS}\right)$	$PRE\left(\hat{\pi}_{HS(2)}, \hat{\pi}_{GS}\right)$	$PRE\left(\hat{\pi}_{HS(3)}, \hat{\pi}_{GS}\right)$
0.1	101.31	121.74	130.67
0.2	100.87	118.69	121.04
0.3	100.71	118.65	118.30
0.4	100.64	119.90	117.70
0.5	100.62	122.23	118.33
0.6	100.63	125.93	120.07
0.7	100.68	131.88	123.27
0.8	100.78	142.27	128.99
0.9	100.96	164.23	140.46

It is observed from Table 5.2 that the percent relative efficiencies of the proposed estimators $\hat{\pi}_{HS(1)}$, $\hat{\pi}_{HS(2)}$ and $\hat{\pi}_{HS(3)}$ with respect to Gjestvang and Singh's (2006) estimator $\hat{\pi}_{GS}$ are larger than 100. It follows that the proposed estimators are more efficient than Gjestvang and Singh's (2006) estimator $\hat{\pi}_{GS}$. We note that there is a marginal gain in efficiency by using the proposed estimator $\hat{\pi}_{HS(1)}$ over Gjestvang and Singh's (2006) estimator $\hat{\pi}_{GS}$ while the gain in efficiency is substantial by using the suggested estimators $\hat{\pi}_{HS(2)}$ and $\hat{\pi}_{HS(3)}$. The proposed estimator $\hat{\pi}_{HS(2)}$ is more efficient than the estimator $\hat{\pi}_{HS(3)}$ as long as $\pi < 0.3$. On the other hand, if $\pi \geq 0.3$ the proposed estimator $\hat{\pi}_{HS(3)}$ is better than the estimator $\hat{\pi}_{HS(2)}$. However, the proposed estimators $\hat{\pi}_{HS(2)}$ and $\hat{\pi}_{HS(3)}$ are more efficient than the estimator $\hat{\pi}_{HS(1)}$. Thus, we conclude that the proposed estimator $\hat{\pi}_{HS(2)}$ is a suitable choice for $\pi < 0.3$, whereas for $\pi \geq 0.3$, the estimator $\hat{\pi}_{HS(3)}$ is the appropriate choice for estimating the population proportion $\hat{\pi}_{HS(3)}$.

Finally, we conclude that the suggested general procedure is justifiable in the sense of obtaining better estimators from the proposed generalized estimator $\hat{\pi}_{HS}$ for appropriate values of (w_1, w_2).

REFERENCES

ANTONAK, R. F., LIVNEH, H., (1995). Randomized response technique: A review and proposed extension to disability attitude research. Genetic, Social, and general Psychology Monographs, 121, pp. 97–145.

BOUZA, C. N., HERRERA, C., MITRA, P. G., (2010). A Review of Randomized Responses Procedures: The Qualitative Variable Case. Revista Investigación Operacional, 31 (3), 240–247.

CHAUDHURI, A., MUKERJEE, R., (1987). Randomized Response Technique: A Review. Statistica Neerlandica, 41, pp. 27–44.

CHAUDHURI, A., MUKERJEE, R., (1988). Randomized Response. Statistics: Textbooks and Monographs, Vol. 85, Marcel Dekker, Inc., New York, NY.

CHAUDHURI, A., CHRISTOFIDES, T. C., RAO, C. R., (2016). Data Gathering, Analysis and Protection of Privacy Through Randomized Response Techniques: Qualitative and Quantitative Human Traits. Handbook of Statistics 34.

FISHER, M., KUPFERMAN, L. B., LESSER, M., (1992). Substance use in a school-based clinic population: Use of the randomized response technique to estimate prevalence. Journal of Adolescent Health, 13, pp. 281–285.

GJESTVANG, C. R., SINGH, S., (2006). A New Randomized Response Model. Journal of the Royal Statistical Society, B. 68, pp. 523–530.

JARMAN, B. J., (1997). The Prevalence and Precedence of Socially Condoned Sexual Aggression Within a Dating Context as Measured by Direct Questioning and the Randomized Response Technique

MANGAT, N. S., SINGH, R., (1990). An Alternative Randomized Procedure. Biometrika, 77, pp. 439–442.

MANGAT, N. S., (1994). An Improved Randomized Response Strategy. Journal of the Royal Statistical Society., B, 56 (1), pp. 93–95.

ODUMADE, O., SINGH, S., (2008). Generalized Force Quantitative Randomized Response Model: A Unified Approch. Journal of the Indian Society and Agricultural Statistics, 62 (3), pp. 244–252.

ODUMADE, O., SINGH, S., (2009). Improved Bar-Lev. Bobovitch, Boukai Randomized Response Model. Communications in Statistics – Simulation and Computation, 38, pp. 473–502.

ODUMADE, O., SINGH, S., (2009). Efficient Use of Two Decks of Cards in Randomized Response Sampling. Communications in Statistics – Theory and Methods 38, pp. 439–446.

ODUMADE, O., SINGH, S., (2010). An Alternative to The Bar-Lev, Bobovitch and Boukai Randomized Response Model. Sociological Methods And Research. Doi10.1177/0049124110378094.

SINGH, S., (2003). Advanced Sampling Theory With Applications. Kluwer Academic Publishers, Dordrecht.

WARNER, S. L., (1965). Randomized Response: A Survey Technique For Eliminating Evasive Answer Bias. Journal of the American Statistical Association, 60, pp. 63–69.

WEISSMAN, A. N., STEER, R. A., LIPTON, D. S., (1986). Estimating illict drug use through telephone interviews and the randomized response technique. Drug and Alcohol Dependence, 18, 225–233.

WILLIAMS, B. L., SUEN, H., (1994). A Methodological Comparison of Survey Techniques in Obtaining Self-Reports of Condom-Related Behaviors. Psychological Reports, 7, pp. 1531–1537.