Statistics and Applications {ISSN 2452-7395 (online)}
Volume 15 Nos. 1&2, 2017 (New Series), pp 7-18

Generalized Multi-Stage Optional Unrelated Question RRT Models

B.K. Dass and Anu Chhabra
Department of Mathematics, University of Delhi, Delhi

Received: December 23, 2016 ; Reviewed: January 05, 2017; Accepted: February 05, 2017

Abstract

We propose a generalized version of Chhabra et al. (2016) multi-stage optional unrelated question RRT model both for binary and quantitative response situations, wherein the prevalence of sensitive variable and the sensitivity level of the underlying sensitive question are estimated simultaneously using a two question approach. In the proposed model we have assumed the prevalence of non-sensitive characteristics also unknown. Simulation results indicate that the proposed models are more efficient than the corresponding one-stage model.

Keywords: Optional Unrelated Question Randomized Response Models; Simulation Study; Parameter Estimation; Sensitive Characteristics.

1. Introduction

Surveys on highly personal and stigmatizing questions concerning issues such as drug addiction, abortion, sexual behavior etc. can lead to unreliable data when people interviewed are required to answer directly. The development of an ingenious tool known as randomized response models by Warner (1965) for personal interview surveys has attracted a greater co-operation to collect data on sensitive topics while ensuring respondent anonymity. Since then, have developed several other randomized response models for collecting data on both the qualitative and quantitative variables.

Instead of requesting the respondent to reply affirmatively or negatively to the sensitive research question as in Warner (1965), Greenberg et al. (1969, 1971) stated the alternatives differently. In this model the respondent face the randomizing device in which the sensitive question is asked with known probability p and an innocuous question is asked with probability$(1 - \mathrm{p})$. A question may be perceived as sensitive by one subject and not by the other. Taking this into account, Gupta et al. (2002) introduced optionality in RRT models. In these models, the respondents for whom the research question is not personally sensitive are instructed to directly answer the research question. Rest of the respondent are asked to provide scrambled responses. This idea was integrated with the unrelated question models in Gupta et al. (2013).

In optional RRT models, there are two parameters of interest-the sensitivity level of the question (proportion of respondents in the population who consider the question sensitive)

Corresponding author: Anu Chhabra
E-mail: a.chhabra02@gmail.com

and prevalence of the sensitive characteristic in the population. Estimating these two parameters simultaneously often requires a split-sample approach, and hence a much larger total sample size. Sihm et al. (2016) proposed modified optional unrelated question models- a binary and a quantitative RRT model which offer the respondents the option of answering the sensitive question directly if they find the research question non-sensitive. In their approach, the sensitivity level is estimated first by using the usual Greenberg et al. (1969) model and then the prevalence of the sensitive characteristic is estimated from the same sample by using the Gupta et al. (2013) optional RRT model. Thus both parameters are estimated sequentially using the same sample but asking two questions from each respondent.

Chhabra et al. (2016) proposed three-stage optional unrelated question RRT models both for binary and quantitative response situation. In these models a randomly selected proportion T of respondents answer the main research question truthfully and a known proportion F of respondents provide a randomized response to the research question using Greenberg et al. (1969) model or Greenberg et al. (1971) model (depending upon whether the response is binary or quantitative). The remaining respondents use the Gupta et al. (2013) model. The sensitivity level is estimated from the same sample by using the Greenberg et al. (1969) model and prevalence of the sensitive attribute using Gupta et al. (2013) model. In our current work we propose a generalized version of Chhabra et al. (2016) model by assuming the prevalence of unrelated characteristics while asking Question 1 and Question 2 to be unknown.

In section 2, the theoretical framework for the generalized multi-stage unrelated question RRT models is discussed. A simulation study is presented in Section 3 and it helps validate the findings of Section 2. Some concluding remarks are made in Section 4.

2. The Proposed Model

In this section, we propose a generalized multi-stage unrelated question RRT models- one for binary response situations and one for quantitative response situations, which are generalizations of Chhabra et al. (2016) model.

2.1 Binary Response Situation

In the proposed binary model, all the respondents are asked two questions. The question about sensitivity is asked first via randomization device 1. In this randomization device, the sensitive question is “Do you consider the research question too sensitive for a direct answer?” prevalence of the corresponding unrelated characteristic is assumed unknown. The underlying sensitivity level w and its variance can be estimated from the sample using Greenberg et al. (1969) model. Then, in the model all the respondents of the same sample are asked another question to ascertain the prevalence of the sensitive characteristic in the population using randomization device 2.

In randomization device 2, the same sample is used. A known proportion T of respondents answer the research question truthfully and a known proportion F of respondents provide a randomized response using the Greenberg et al. (1969) model in which the respondent uses the randomization device which bears sensitive question with the known probability p_b and an unrelated innocuous question whose prevalence is unknown which has no possible embarrassment, is answered with probability $(1 - p_b)$. The remaining proportion $(1 - T - F)$ of respondents uses Gupta et al. (2013) optional unrelated question model, in which the respondent is given the option to answer the research question directly (or using

the Greenberg et al. (1969) model with known parameter p_b) if they find the research question non−sensitive (or sensitive).
We use the following notation. Let

- n be the sample size,
- π_a *be* the unknown probability of an unrelated question used in Greenberg et al. (1969) model of randomization device 1 while asking Question 1,
- π_b be the unknown probability of another unrelated question used in Greenberg et al. (1969) model of randomization device 2 while asking Question 2,
- π be the unknown proportion of population that belongs to the sensitive group,
- p_a be the known probability of the respondent selecting the question about sensitivity in Greenberg et al. (1969) model of randomization device 1 while asking Question 1,
- p_b be the known probability of the respondent selecting the sensitive question in Greenberg et al. (1969) model of randomization device 2 while asking Question 2,
- w be the sensitivity level of the main research question in the population,
- P_{Y1} be the probability of 'yes' response from a respondent in randomization device 1 while asking Question 1,
- P_{Y2} be the probability of 'yes' response from a respondent in randomization device 2 while asking Question 2.

Thus, using randomization device 1, we obtain

$$P_{Y1} = p_a w + (1 - p_a)\pi_a \tag{1}$$

Since there are two unknowns w and π_a we need two equations. So we split the whole sample of size n in two sub-samples of sizes n_1 and n_2. Different scrambling devices are used in both the sub-samples.

Let

- p_{a1} be the known probability of the respondent selecting the question about sensitivity in Greenberg et al. (1969) model of randomization device 1 while asking Question 1 from a sub-sample of size n_1,
- p_{a2} be the known probability of the respondent selecting the question about sensitivity in Greenberg et al. (1969) model of randomization device 1 while asking Question 1 from a sub-sample of size n_2,
- P_{Y11} be the probability of 'yes' response by a respondent in randomization device 1 while asking Question 1 from a sub-sample of size n_1,
- P_{Y12} be the probability of 'yes' response by a respondent in randomization device 1 while asking Question 1 from a sub-sample of size n_2.

Then,

$$P_{Y11} = p_{a1} w + (1 - p_{a1})\pi_a \tag{2}$$

$$P_{Y12} = p_{a2} w + (1 - p_{a2})\pi_a \tag{3}$$

Solving for w ,

$$w = \frac{(1 - p_{a2})P_{Y11} - (1 - p_{a1})P_{Y12}}{p_{a1} - p_{a2}} \tag{4}$$

Solving for π_a, $\quad \pi_a = \dfrac{p_{a2}P_{Y11} - p_{a1}P_{Y12}}{p_{a2} - p_{a1}}$ (5)

Thus, the estimate of w is given by, $\hat{w} = \dfrac{(1-p_{a2})\hat{P}_{Y11} - (1-p_{a1})\hat{P}_{Y12}}{p_{a1} - p_{a2}}$ (6)

where $\hat{P}_{Y11}$ is the proportion of "yes" responses while asking Question 1 from a sample of size $n_1$and $\hat{P}_{Y12}$ be the proportion of "yes" responses while asking Question 1 from a sample of size n_2.

After applying first order Taylor's Approximation to equation (6) we have,

$$\hat{w} = \hat{w}(P_{Y11}, P_{Y12}) + \left.\frac{\partial \hat{w}(\hat{P}_{Y11}, \hat{P}_{Y12})}{\partial \hat{P}_{Y11}}\right|_{(P_{Y11}, P_{Y12})} (\hat{P}_{Y11} - P_{Y11}) + \left.\frac{\partial \hat{w}(\hat{P}_{Y11}, \hat{P}_{Y12})}{\partial \hat{P}_{Y12}}\right|_{(P_{Y11}, P_{Y12})} (\hat{P}_{Y12} - P_{Y12}) \quad (7)$$

i.e. $$\hat{w} = \frac{(1-p_{a2})P_{Y11} - (1-p_{a1})P_{Y12}}{p_{a1} - p_{a2}} + \frac{(1-p_{a2})}{(p_{a1}-p_{a2})}(\hat{P}_{Y11} - P_{Y11}) - \frac{(1-p_{a1})}{(p_{a1}-p_{a2})}(\hat{P}_{Y12} - P_{Y12}) \quad (8)$$

The expectation of $\hat{w}$, to the first order of Taylor's Approximation, is given by

$$E(\hat{w}) = \frac{(1-p_{a2})P_{Y11} - (1-p_{a1})P_{Y12}}{p_{a1} - p_{a2}} = w. \quad (9)$$

and $$Var(\hat{w}) = \frac{(1-p_{a2})^2}{(p_{a1}-p_{a2})^2}\frac{P_{Y11}(1-P_{Y11})}{n_1} + \frac{(1-p_{a1})^2}{(p_{a1}-p_{a2})^2}\frac{P_{Y12}(1-P_{Y12})}{n_2}. \quad (10)$$

For Question 2, let P_{Y2} be the probability of 'yes' response from a respondent to question 2. We have,

$$P_{Y2} = T\pi + F\{p_b\pi + (1-p_b)\pi_b\} + (1-T-F)\{(1-w)\pi + w[p_b\pi + (1-p_b)\pi_b]\}. \quad (11)$$

Solving, the above equation we have

$$P_{Y2} = \{(1-w)(1-F(1-p_b)) + w(p_b + T(1-p_b))\}\pi + (1-p_b)\{F + w(1-T-F)\}\pi_b \quad (12)$$

Since there are two unknowns π and π_b we need two equations. So we split the whole sample of size n in two sub-samples of sizes n_1 and n_2. Different scrambling devices are used in both the sub-samples. Let

- p_{b1} be the known probability of the respondent selecting the sensitive question in Greenberg et al. (1969) model of randomization device 2 while asking Question 2 from a sub-sample of size n_1,
- p_{b2} be the known probability of the respondent selecting the sensitive question in Greenberg et al. (1969) model of randomization device 2 while asking Question 2 from a sub-sample of size n_2,
- P_{Y21} be the probability of 'yes' response from a respondent in randomization device 2 while asking Question 2 from the same sub-sample of size n_1,
- P_{Y22} be the probability of 'yes' response from a respondent in randomization device 2 while asking Question 2 from the same sub-sample of size n_2.

Thus,

$$P_{Y21} = \{(1-w)(1-F(1-p_{b1})) + w(p_{b1} + T(1-p_{b1}))\}\pi + (1-p_{b1})\pi_b\{F + w(1-T-F)\}. \quad (13)$$

$$P_{Y22} = \{(1-w)(1-F(1-p_{b2})) + w(p_{b2} + T(1-p_{b2}))\}\pi + (1-p_{b2})\pi_b\{F + w(1-T-F)\}. \quad (14)$$

Solving equation (13) and equation (14) for π_b we have,

$$\pi_b = \frac{\{(1-w)(1-F(1-p_{b2})) + w(p_{b2} + T(1-p_{b2}))\}P_{Y21} - \{(1-w)(1-F(1-p_{b1})) + w(p_{b1} + T(1-p_{b1}))\}P_{Y22}}{\{(1-w)(1-F(1-p_{b2})) + w(p_{b2} + T(1-p_{b2}))\}(1-p_{b1})(F + w(1-T-F)) - \{(1-w)(1-F(1-p_{b1})) + w(p_{b1} + T(1-p_{b1}))\}(1-p_{b2})(F + w(1-T-F))}. \quad (15)$$

Let,

$$\lambda = \frac{\{(1-w)(1-F(1-p_{b2})) + w(p_{b2} + T(1-p_{b2}))\}}{\{(1-w)(1-F(1-p_{b1})) + w(p_{b1} + T(1-p_{b1}))\}}. \quad (16)$$

$$\pi_b = \frac{\lambda P_{Y21} - P_{Y22}}{\{F + w(1-T-F)\}\{\lambda(1-p_{b1}) - (1-p_{b2})\}}. \quad (17)$$

Solving equation (13) and equation (14) for π, we have

$$\pi = \frac{(1-p_{b2})P_{Y21} - (1-p_{b1})P_{Y22}}{\{(1-w)(1-F(1-p_{b1})) + w(p_{b1} + T(1-p_{b1}))\}(1-p_{b2}) - \{(1-w)(1-F(1-p_{b2})) + w(p_{b2} + T(1-p_{b2}))\}(1-p_{b1})} \quad (18)$$

$$\pi = \frac{(1-p_{b2})P_{Y21} - (1-p_{b1})P_{Y22}}{(p_{b1} - p_{b2})} \quad (19)$$

Thus, the estimate of π is given by

$$\hat{\pi} = \frac{(1-p_{b2})\hat{P}_{Y21} - (1-p_{b1})\hat{P}_{Y22}}{(p_{b1} - p_{b2})}. \quad (20)$$

After applying first order Taylor's Approximation to equation (20) we have,

$$\hat{\pi} = \hat{\pi}(P_{Y21}, P_{Y22}) + \left.\frac{\partial\hat{\pi}(\hat{P}_{Y21}, \hat{P}_{Y22})}{\partial\hat{P}_{Y21}}\right|_{(P_{Y21}, P_{Y22})}(\hat{P}_{Y21} - P_{Y21}) + \left.\frac{\partial\hat{\pi}(\hat{P}_{Y21}, \hat{P}_{Y22})}{\partial\hat{P}_{Y22}}\right|_{(P_{Y21}, P_{Y22})}(\hat{P}_{Y22} - P_{Y22}). \quad (21)$$

$$\hat{\pi} = \frac{(1-p_{b2})P_{Y21} - (1-p_{b1})P_{Y22}}{(p_{b1} - p_{b2})} + \frac{(1-p_{b2})}{(p_{b1} - p_{b2})}(\hat{P}_{Y21} - P_{Y21}) - \frac{(1-p_{b1})}{(p_{b1} - p_{b2})}(\hat{P}_{Y22} - P_{Y22}). \quad (22)$$

The expectation of $\hat{\pi}$, to the first order of Taylor's Approximation, is given by

$$E(\hat{\pi}) = \frac{(1-p_{b2})P_{Y21} - (1-p_{b1})P_{Y22}}{(p_{b1} - p_{b2})} = \pi. \quad (23)$$

and

$$\mathrm{Var}(\hat{\pi}) = \frac{(1-p_{b2})^2}{(p_{b1} - p_{b2})^2}\frac{P_{Y21}(1-P_{Y21})}{n_1} + \frac{(1-p_{b1})^2}{(p_{b1} - p_{b2})^2}\frac{P_{Y22}(1-P_{Y22})}{n_2}. \quad (24)$$

2.2 Quantitative Response Situation

In the proposed quantitative model, all the respondents are asked two questions. The question about sensitivity is asked first using randomization device 1. In this randomization device, the sensitive question is "Do you consider the research question too sensitive for a direct answer?" prevalence of the corresponding unrelated characteristic is assumed unknown. The underlying sensitivity level w and its variance can be estimated from the sample using Greenberg et al. (1969) model. All the respondents of the same sample are asked another question to ascertain the mean prevalence of the sensitive characteristic X in the population. This is done using randomization device 2 using the same sample as the one used for the first question. With this randomization device, a known proportion T of respondents answer the research question truthfully and a known proportion F of respondents provide a randomized response using Greenberg et al. (1971) model in which the respondent uses the randomization device which bears sensitive question with the known probability p_b and an unrelated innocuous question is answered with probability $(1-p_b)$. The remaining proportion $(1-T-F)$ of respondents use Gupta et al. (2013) optional unrelated question model, in which the respondent is given the option to answer the research question directly (or by using the Greenberg et al. (1971) model with known parameter p_b and unknown mean and variance of the innocuous variable) if they find the research question non−sensitive (sensitive).

We use the following notation:

- n be the sample size,
- w be the sensitivity level of the survey question in the population,
- π_a be the unknown probability of an unrelated question used in Greenberg et al. (1969) model of randomization device 1 while asking Question 1,
- p_a be the known probability of the respondent selecting the question about sensitivity in Greenberg et al. (1969) model of randomization device 1 while asking Question 1,
- p_b be the known probability of the respondent selecting the sensitive question in Greenberg et al. (1971) model of randomization device 2 while asking Question 2,
- μ_Y and σ_Y^2 be the unknown mean and variance of an innocuous question used in Greenberg et al. (1971) model of randomization device 2 while asking Question 2,
- μ_X and σ_X^2 be the unknown mean and variance of the sensitive question of the population,
- Z be the reported quantitative response to randomization device 2 while asking Question 2 by a respondent.

Then from randomization device 1, we obtain $\hat{w}$ as an unbiased estimator of w and $\hat{w}$ with its variance are given in equation (8) and equation (10) above respectively.

From randomization device 2, we get

$$Z = \begin{cases} X \\ Y \end{cases} \text{with probability} \begin{cases} T + Fp_b + (1-T-F)\left[(1-w) + wp_b\right] & \text{(sensitive question)} \\ F\left(1-p_b\right) + (1-T-F)w(1-p_b) & \text{(innocuous question)} \end{cases}$$

The expectation and variance of Z is given by,

$$E(Z) = E(X)\{T + \left[Fp_b + (1-T-F)(1-w+wp_b)\right]\} + E(Y)\left[F(1-p_b) + (1-T-F)w(1-p_b)\right] \tag{25}$$

and

$$\text{Var}(Z) = \{T + \left[Fp_b + (1-T-F)(1-w+wp_b)\right]\}E(X^2) + \left[F(1-p_b) + (1-T-F)w(1-p_b)\right]E(Y^2) - \mu_Z^2 \tag{26}$$

$$\text{Var}(Z) = \{T + \left[Fp_b + (1-T-F)(1-w+wp_b)\right]\}\left(\sigma_X^2 + \mu_X^2\right) +$$

$$\left[F(1-p_b)+(1-T-F)w(1-p_b)\right]\left(\sigma_Y^2+\mu_Y^2\right)-\mu_Z^2 . \tag{27}$$

Since, there are two unknowns μ_X and μ_Y we need to follow split-sample approach. We use the same sample split as done in Question 1.

Let

- p_{b1} be the known probability of the respondent selecting the sensitive question in Greenberg et al. (1971) model of randomization device 2 while asking Question 2 from a sub-sample of size n_1,
- p_{b2} be the known probability of the respondent selecting the sensitive question in Greenberg et al. (1971) model of randomization device 2 while asking Question 2 from a sub-sample of size n_2,
- Z_1 be the reported response to randomization device 2 by a respondent while asking Question 2 from a sub-sample of size n_1,
- Z_2 be the reported response to randomization device 2 by a respondent while asking Question 2 from a sub-sample of size n_2.

We have,

$$E(Z_1)=E(X)\{T+Fp_{b1}+(1-T-F)(1-w+wp_{b1})\}+E(Y)(1-p_{b1})\left[F+(1-T-F)w\right] \tag{28}$$

i.e.

$$\mu_{Z1}=\mu_X\{T+Fp_{b1}+(1-T-F)(1-w+wp_{b1})\}+\mu_Y(1-p_{b1})\left[F+(1-T-F)w\right] \tag{29}$$

and

$$E(Z_2)=E(X)\{T+Fp_{b2}+(1-T-F)(1-w+wp_{b2})\}+E(Y)(1-p_{b2})\left[F+(1-T-F)w\right] \tag{30}$$

i.e.

$$\mu_{Z2}=\mu_X\{T+Fp_{b2}+(1-T-F)(1-w+wp_{b2})\}+\mu_Y(1-p_{b2})\left[F+(1-T-F)w\right] \tag{31}$$

Solving equation (29) and equation (31) for μ_X, we have

$$\mu_X=\frac{(1-p_{b2})\mu_{Z1}-(1-p_{b1})\mu_{Z2}}{p_{b1}-p_{b2}} \tag{32}$$

Thus, the estimate of μ_X is given by

$$\hat{\mu}_X=\frac{(1-p_{b2})\hat{\mu}_{Z1}-(1-p_{b1})\hat{\mu}_{Z2}}{p_{b1}-p_{b2}} \tag{33}$$

After applying first order Taylor's Approximation to equation (33) we have,

$$\hat{\mu}_X=\hat{\mu}_X(\mu_{Z1},\mu_{Z2})+\left.\frac{\partial\hat{\mu}_X(\hat{\mu}_{Z1},\hat{\mu}_{Z2})}{\partial\hat{\mu}_{Z1}}\right|_{(\mu_{Z1},\mu_{Z2})}(\hat{\mu}_{Z1}-\mu_{Z1})+\left.\frac{\partial\hat{\mu}_X(\hat{\mu}_{Z1},\hat{\mu}_{Z2})}{\partial\hat{\mu}_{Z2}}\right|_{(\mu_{Z1},\mu_{Z2})}(\hat{\mu}_{Z2}-\mu_{Z2}) \tag{34}$$

i.e.

$$\hat{\mu}_X=\frac{(1-p_{b2})\mu_{Z1}-(1-p_{b1})\mu_{Z2}}{p_{b1}-p_{b2}}+\frac{(1-p_{b2})}{(p_{b1}-p_{b2})}(\hat{\mu}_{Z1}-\mu_{Z1})+\frac{(-1)(1-p_{b1})}{(p_{b1}-p_{b2})}(\hat{\mu}_{Z2}-\mu_{Z2}) \tag{35}$$

The expectation of $\hat{\mu}_X$, to the first order of Taylor's Approximation, is given by

$$E(\hat{\mu}_X) = \frac{(1-p_{b2})\mu_{Z1} - (1-p_{b1})\mu_{Z2}}{p_{b1} - p_{b2}} = \mu_X . \tag{36}$$

and

$$\mathrm{Var}(\hat{\mu}_X) = \frac{(1-p_{b2})^2}{(p_{b1}-p_{b2})^2}\mathrm{Var}(\hat{\mu}_{Z1}) + \frac{(1-p_{b1})^2}{(p_{b1}-p_{b2})^2}\mathrm{Var}(\hat{\mu}_{Z2}) . \tag{37}$$

where,

$$\mathrm{Var}(\hat{\mu}_{Z1}) = \frac{\mathrm{Var}(Z_1)}{n_1} . \tag{38}$$

and
$$\mathrm{Var}\ (\hat{\mu}_{Z2}) = \frac{\mathrm{Var}\ (Z_2)}{n_2} . \tag{39}$$

Solving equation (29) and equation (31) for μ_Y, we have

$$\mu_Y = \frac{\{T + Fp_{b2} + (1-T-F)(1-w+wp_{b2})\}\mu_{Z1} - \{T + Fp_{b1} + (1-T-F)(1-w+wp_{b1})\}\mu_{Z2}}{(1-p_{b1})[F + (1-T-F)w]\{T + Fp_{b2} + (1-T-F)(1-w+wp_{b2})\} - (1-p_{b2})[F + (1-T-F)w]\{T + Fp_{b1} + (1-T-F)(1-w+wp_{b1})\}} . \tag{40}$$

Let

$$\lambda = \frac{\{T + Fp_{b2} + (1-T-F)(1-w+wp_{b2})\}}{\{T + Fp_{b1} + (1-T-F)(1-w+wp_{b1})\}} \tag{41}$$

So,

$$\mu_Y = \frac{\lambda\mu_{Z1} - \mu_{Z2}}{\{(1-p_{b1})\lambda - (1-p_{b2})\}[F + (1-T-F)w]} \tag{42}$$

3. Simulation Results

In this section, the theoretical results obtained in Section 2 for our estimators $\hat{\mu}_X$, $\hat{\pi}$ and $\hat{w}$ are verified empirically. All the simulations were conducted using SAS programming language. For the binary response models, parameters T and F , were allowed to vary while all other variables were fixed. We used number of trials = 10000, $w = 0.9$, $\pi = 0.3$, $\pi_a = 0.1$, $\pi_b = 0.7$, $p_{a1} = 0.2$, $p_{a2} = 0.7$, $p_{b1} = 0.3$, $p_{b2} = 0.9$ and $n = 1000$. The whole sample is equally divided in two sub-samples of sizes $n_1 = 500$ and $n_2 = 500$. For the quantitative response models, parameters T and F were allowed to vary while all other variables were fixed again and we used the number of trials = 10000, $w = 0.9$, $\pi_a = 0.1$, $p_{a1} = 0.2$, $p_{a2} = 0.7$, $p_{b1} = 0.3$, $p_{b2} = 0.9$, $\mu_X = 2$, $\mu_Y = 7$ and $n = 1000$. The whole sample is equally divided in two sub-samples of sizes $n_1 = 500$ and $n_2 = 500$. Further, X and Y are assumed to follow Poisson distribution with parameters μ_X and μ_Y respectively. Both the proposed models are valid for those combinations of T and F for which $T + F < 1$. Thus, in tables given below, the combinations of T and F for which $T + F \geq 1$ are marked with a dash $(-)$.

3.1 Simulation of $\hat{\pi}$ and $\hat{w}$ for generalized binary three-stage model

The table below shows the simulation results calculated using SAS programming language for different combination of T and F and the theoretical values are calculated using the results obtained in Section 2.

Table 1: Simulation Results of Binary Model

Trials =10000, $w = 0.9$, $\pi = 0.3$, $\pi_a = 0.1$, $\pi_b = 0.7$, $p_{a1} = 0.2$, $p_{a2} = 0.7$, $p_{b1} = 0.3$, $p_{b2} = 0.9$, $n_1 = 500$, $n_2 = 500$.

F		T				
		0	0.1	0.3	0.5	0.7
0	Sim. Var $(\hat{w})$	0.0012883	0.0012722	0.0012722	0.0012722	0.0012722
	Theo. Var $(\hat{w})$	0.0012	0.0012	0.0012	0.0012	0.0012
	Sim. Mean $(\hat{w})$	0.9002187	0.9001922	0.9001922	0.9001922	0.9001922
	Sim. Var $(\hat{\pi})$	0.000622665	0.000627936	0.000621135	0.000611242	0.000592475
	Theo. Var $(\hat{\pi})$	0.000621077	0.000617938	0.000611236	0.000603969	0.000596138
	Sim. Mean $(\hat{\pi})$	0.2998357	0.3003454	0.3004091	0.3004734	0.3003022
	Sim. Var $(\hat{\pi})$ / Theo. Var $(\hat{\pi})$	1.002556849	1.016179617	1.016195054	1.012042009	0.99385545
0.1	Sim. Var $(\hat{w})$	0.0012722	0.0012732	0.0012732	0.0012732	0.0012732
	Theo. Var $(\hat{w})$	0.0012	0.0012	0.0012	0.0012	0.0012
	Sim. Mean $(\hat{w})$	0.9001922	0.8998585	0.8998585	0.8998585	0.8998585
	Sim. Var $(\hat{\pi})$	0.000634551	0.000617069	0.000611192	0.000603195	0.000593805
	Theo. Var $(\hat{\pi})$	0.000621417	0.000618294	0.000611623	0.000604388	0.000596588
	Sim. Mean $(\hat{\pi})$	0.300266	0.2998498	0.2999399	0.2999252	0.2999976
	Sim. Var $(\hat{\pi})$ /Theo. Var $(\hat{\pi})$	1.021135566	0.998018742	0.999295318	0.998026102	0.995335139
0.3	Sim. Var $(\hat{w})$	0.0012722	0.0012732	0.0012732	0.0012732	-
	Theo. Var $(\hat{w})$	0.0012	0.0012	0.0012	0.0012	-
	Sim. Mean $(\hat{w})$	0.8998585	0.8998585	0.8998585	0.8998585	-
	Sim. Var $(\hat{\pi})$	0.000635883	0.000617936	0.000612398	0.000604051	-
	Theo. Var $(\hat{\pi})$	0.000622092	0.000619	0.000612392	0.00060522	-
	Sim. Mean $(\hat{\pi})$	0.3002504	0.2998577	0.2999395	0.2999288	-
	Sim. Var $(\hat{\pi})$ / Theo. Var $(\hat{\pi})$	1.022168747	0.998281099	1.000009798	0.998068471	-
0.5	Sim. Var $(\hat{w})$	0.0012722	0.0012732	0.0012732	-	-
	Theo. Var $(\hat{w})$	0.0012	0.0012	0.0012	-	-
	Sim. Mean $(\hat{w})$	0.9001922	0.8998585	0.8998585	-	-
	Sim. Var $(\hat{\pi})$	0.000636683	0.000618112	0.000612135	-	-
	Theo. Var $(\hat{\pi})$	0.00062276	0.000619699	0.000613154	-	-
	Sim. Mean $(\hat{\pi})$	0.3002484	0.2998882	0.299963	-	-
	Sim. Var $(\hat{\pi})$ / Theo. Var $(\hat{\pi})$	1.022356927	0.997439079	0.998338101	-	-
0.7	Sim. Var $(\hat{w})$	0.0012722	0.0012732	-	-	-
	Theo. Var $(\hat{w})$	0.0012	0.0012	-	-	-
	Sim. Mean $(\hat{w})$	0.9001922	0.8998585	-	-	-
	Sim. Var $(\hat{\pi})$	0.000634322	0.000617312	-	-	-
	Theo. Var $(\hat{\pi})$	0.000623421	0.000620392	-	-	-
	Sim. Mean $(\hat{\pi})$	0.3002211	0.299885	-	-	-
	Sim. Var $(\hat{\pi})$ / Theo. Var $(\hat{\pi})$	1.017485776	0.995035397	-	-	-

The simulation results support our theoretical finding that $\hat{\pi}$ and $\hat{w}$ are unbiased estimators of π and w respectively up to first order Approximation. The theoretical and simulated variances of $\hat{w}$ are very close and the theoretical and simulated variances of $\hat{\pi}$ are also very close as indicated by the ratios in the last row of each block. For example, one may observe from Table 1 that for $T = 0.3$ and $F = 0.3$, the theoretical value of $Var(\hat{w}) = 0.0012$ and simulated value of $Var(\hat{w}) = 0.0012732$. Similarly, for $T = 0.3$ and $F = 0.3$, the theoretical value of $Var(\hat{\pi}) = 0.000612392$ and simulated value of $Var(\hat{\pi}) = 0.000612398$. The first order Taylor's Approximation was used to calculate the theoretical values for $Var(\hat{w})$ and $Var(\hat{\pi})$.

3.2. Simulation of $\hat{\mu}_X$ and $\hat{w}$ for generalized quantitative three-stage model

The table below shows the simulation results calculated using SAS programming language for different combination of T and F and the theoretical values are calculated using the results obtained in Section 2.

Table 2 : Simulation Results of Quantitative Model

Trials = 10000, $w = 0.9$, $\mu_X = 2$, $\mu_Y = 7$, $\pi_a = 0.1$, $p_{a1} = 0.2$, $p_{a2} = 0.7$, $p_{b1} = 0.3$, $p_{b2} = 0.9$, $n = 1000$, $n_1 = 500$, $n_2 = 500$.

F			T				
			0	0.1	0.3	0.5	0.7
0		Sim. Var $(\hat{w})$	0.0012869	0.0012869	0.0012869	0.0012869	0.0012869
		Theo. Var $(\hat{w})$	0.0012	0.0012	0.0012	0.0012	0.0012
		Sim. Mean $(\hat{w})$	0.8999578	0.8999578	0.8999578	0.8999578	0.8999578
		Sim. Var $(\hat{\mu}_X)$	0.0127931	0.0122525	0.010834	0.0096872	0.0081281
		Theo. Var $(\hat{\mu}_X)$	0.012853056	0.012222531	0.010895331	0.009479931	0.007976331
		Sim.Mean $(\hat{\mu}_X)$	1.9998117	1.9999975	1.9997799	1.9995907	2.0000172
		Sim. Var $(\hat{\mu}_X)$ / Theo. Var $(\hat{\mu}_X)$	0.995335273	1.002451947	0.994370892	1.021863978	1.01902742
0.1	Proposed Model	Sim. Var $(\hat{w})$	0.0012869	0.0012869	0.0012869	0.0012869	0.0012869
		Theo. Var $(\hat{w})$	0.0012	0.0012	0.0012	0.0012	0.0012
		Sim. Mean $(\hat{w})$	0.8999578	0.8999578	0.8999578	0.8999578	0.8999578
		Sim. Var $(\hat{\mu}_X)$	0.0128918	0.0123174	0.0109065	0.0097501	0.0082198
		Theo. Var $(\hat{\mu}_X)$	0.012921753	0.012293678	0.010971378	0.009560878	0.008062178
		Sim.Mean $(\hat{\mu}_X)$	1.999786	1.9999661	1.9996712	1.9996388	1.9999975
		Sim. Var $(\hat{\mu}_X)$ / Theo. Var $(\hat{\mu}_X)$	0.997681971	1.00192961	0.994086613	1.019791279	1.019550796
0.3	Proposed Model	Sim. Var $(\hat{w})$	0.0012869	0.0012869	0.0012869	0.0012869	-
		Theo. Var $(\hat{w})$	0.0012	0.0012	0.0012	0.0012	-
		Sim. Mean $(\hat{w})$	0.8999578	0.8999578	0.8999578	0.8999578	-
		Sim. Var $(\hat{\mu}_X)$	0.0130073	0.0124002	0.0110545	0.0098954	-
		Theo. Var $(\hat{\mu}_X)$	0.013058331	0.012435156	0.011122656	0.009721956	-
		Sim.Mean $(\hat{\mu}_X)$	1.9998812	1.9999859	1.9997499	1.9995702	-

		Sim. Var $(\hat{\mu}_X)$ / Theo. Var $(\hat{\mu}_X)$	0.996092073	0.997188938	0.993872327	1.017840443	-

0.5	Proposed Model	Sim. Var $(\hat{w})$	0.0012869	0.0012869	0.0012869	-	-
		Theo. Var $(\hat{w})$	0.0012	0.0012	0.0012	-	-
		Sim. Mean $(\hat{w})$	0.8999578	0.8999578	0.8999578	-	-
		Sim. Var $(\hat{\mu}_X)$	0.0131078	0.0125434	0.0112264	-	-
		Theo. Var $(\hat{\mu}_X)$	0.013193819	0.012575544	0.011272844	-	-
		Sim.Mean $(\hat{\mu}_X)$	1.9999213	1.9999244	1.9996857	-	-
		Sim. Var $(\hat{\mu}_X)$ / Theo. Var $(\hat{\mu}_X)$	0.993480356	0.997443928	0.99588001	-	-
0.7	Proposed Model	Sim. Var $(\hat{w})$	0.0012869	0.0012869	-	-	-
		Theo. Var $(\hat{w})$	0.0012	0.0012	-	-	-
		Sim. Mean $(\hat{w})$	0.8999578	0.8999578	-	-	-
		Sim. Var $(\hat{\mu}_X)$	0.0132067	0.012656	-	-	-
		Theo. Var $(\hat{\mu}_X)$	0.013328219	0.012714844	-	-	-
		Sim.Mean $(\hat{\mu}_X)$	1.999889	1.9997843	-	-	-
		Sim. Var $(\hat{\mu}_X)$ / Theo. Var $(\hat{\mu}_X)$	0.990882578	0.995372023	-	-	-

In this case also, the simulations results help validate our theoretical findings. It may be noted from Table 2 that unbiasedness of $\hat{\mu}_X$ and $\hat{w}$ is well-supported. The theoretical and simulated variances of $\hat{\mu}_X$ are also very close as indicated by the ratios in the last row of each block. For example, note from Table 2 that for $T = 0.3$ and $F = 0.3$, the theoretical value of $Var(\hat{w}) = 0.0012$ and corresponding simulated variance is $Var(\hat{w}) = 0.0012869$. Similarly, the theoretical value of $Var(\hat{\mu}_X) = 0.011122656$ and the corresponding simulated value is 0.0110545. The first order Taylor's Approximation was used to calculate the theoretical values for $Var(\hat{w})$ and $Var(\hat{\mu}_X)$.

4. Conclusions

In this paper, we propose a generalized version of Chhabra et al. (2016) three-stage modified optional unrelated question RRT models for both binary and quantitative response situations. The simulation study shows that for appropriate choices of (T, F), the proposed models work better than the corresponding one-stage models (i.e. T=0=F). For example: one may observe from Table 1, that for $T = 0.5$ and $F = 0.3$, the theoretical value of $Var(\hat{\pi}) = 0.00060522$ and $T = 0$ and $F = 0$, the theoretical value of $Var(\hat{\pi}) = 0.000621077$. Similarly, one may observe from Table 2, that for $T = 0.5$ and $F = 0.3$, the theoretical value of $Var(\hat{\mu}_X) = 0.009721956$ and for $T = 0$ and $F = 0$, the theoretical value of $Var(\hat{\mu}_X) = 0.012853056$. Also note that $Var(\hat{\mu}_X)$ decreases steadily as T increases. This is because for our choice of various parameters, $Var(\hat{\mu}_X)$ linear function of T with negative coefficient. Similarly for the binary case, $Var(\hat{\pi})$ has its maximum at -2.17 and hence decreases continuously in the 0 to 1 range.

References

Chhabra, A., Dass, B.K. and Mehta S. (2016). Multi-Stage Optional Unrelated Question RRT Model. *Journal of Statistical Theory and Applications*, **15(1)**, 80-95.

Greenberg, B.G., Abul-Ela, A.L.A., Simmons, W.R. and Horvitz, D.G. (1969). The Unrelated Question Randomized Response Model: Theoretical Framework. *Journal of the American Statistical Association*, **64**, 520-529.

Greenberg, B.G., Kuebler, R.R., Abernathy, J.R. and Horvitz, D.G. (1971). Applications of the randomized response technique in obtaining quantitative data. *Journal of the American Statistical Association*, **66 (334)**, 243-250.

Gupta, S., Gupta, B. and Singh, S. (2002). Estimation of Sensitivity Level of Personal Interview Survey Question. *Journal of Statistical Planning and Inference*, **100(2)**, 239-247.

Gupta, S., Tuck, A., Spears Gill, T. and Crowe, M. (2013). Optional Unrelated Question Randomized Response Models. Involve: *A Journal of Mathematics*, **6(4)**, 483-492.

Sihm, J.S., Chhabra, A. and Gupta, S.N. (2014). An Optional Unrelated Question RRT Model. (Forthcoming in Involve: *A Journal of Mathematics*).

Warner, S.L. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *Journal of American Statistical Association*, **60 (309)**, 63-69.