Brazilian Journal of Probability and Statistics
2014, Vol. 28, No. 3, 381–408
DOI: 10.1214/12-BJPS214

Estimation of the proportion of a sensitive attribute based on a two-stage randomized response model with stratified unequal probability sampling

Gi-Sung Lee[a], Ki-Hak Hong[b], Jong-Min Kim[c] and Chang-Kyoon Son[d,*]

[a]*Woosuk University*
[b]*Dongshin University*
[c]*University of Minnesota–Morris*
[d]*Dongguk University*

Abstract. To estimate the proportion of a sensitive attribute of the population that is composed of the number of different sized clusters, we suggest a two-stage randomized response model with unequal probability sampling by using Abdelfatah et al.'s procedure [*Braz. J. Probab. Stat.* **27** (2013) 608–617]. We compute the estimate of the sensitive parameter, its variance, and the variance estimator for both pps sampling and two-stage equal probability sampling. We extend our model to the case of stratified unequal probability sampling and compute them. Finally, we compare the efficiency of the two estimators, one obtained by unequal probability sampling and the other by stratified unequal probability sampling.

1 Introduction

Warner (1965) first suggested an ingenious survey model called a randomized response model (RRM) to procure sensitive information from respondents without disturbing their privacy by using a randomizing device which was composed of two questions, one sensitive and the other nonsensitive:

Do you have a sensitive attribute A? (with probability P_0)

Do you have a nonsensitive attribute $\bar{A}$? (with probability $1 - P_0$)

Mangat–Singh (1990) developed a two-stage randomized response model which required the use of two random devices (R_1, R_2) and showed that his model was more efficient than Warner's model under the condition of $T_0 > \frac{1-2P_0}{1-P_0}$. The random device R_1 consists of two questions, (i) "Do you have a sensitive attribute A? (with probability T_0)" and (ii) "Go to random device R_2 (with probability $1-T_0$)." The random device R_2 has exactly the same structure as Warner's model.

Mangat (1994) developed a randomized response model which reduced the use of the randomizing device from two to one.

Odumade and Singh (2009) suggested the use of two decks of cards in a randomized response model where each of the decks included the two questions used

Key words and phrases. Randomized response model, two-stage model, sensitive attribute, stratified sampling, stratified unequal probability sampling.

Received September 2012; accepted November 2012.

Table 1 *Classification of responses from Deck*(1) *and Deck*(2)

	Responses from Deck(2)	
Responses from Deck(1)	Yes	No
Yes	n_{11}	n_{10}
No	n_{01}	n_{00}

in Warner's model. Each respondent in a simple random sampling with replacement (SRSWR) of n respondents is provided with two decks of cards. Deck(1) includes the two questions, (a) Do you have a sensitive attribute A? (b) Do you have a nonsensitive attribute $\bar{A}$?, with probabilities P and $1 - P$, respectively. Deck(2) includes the two questions as in Deck(1) with probabilities T and $1 - T$, respectively. Each respondent is requested to draw two cards simultaneously, one by one from each deck of cards, read the questions in order and answer "Yes" or "No" accordingly. The responses from the n respondents can be classified into 2×2 contingency table as shown in Table 1.

Abdelfatah et al. (2013) suggested a modified Odumade and Singh (2009) model that improved its efficiency by using Mangat–Singh's (1990) procedure instead of Warner's procedure at each stage.

In this paper, we suggest a two-stage randomized response model, with unequal probability sampling, to estimate the proportion of the sensitive attribute of the population that is composed of the number of different sized clusters by using Abdelfatah et al.'s procedure (2013). We compute the estimate of the proportion of a sensitive parameter, its variance and variance estimator for both *pps* sampling and two-stage equal probability sampling, respectively. We extend our model to the case of stratified unequal probability sampling and compute them. Finally, we compare the efficiency of the two estimators, one obtained by unequal probability sampling and the other by stratified unequal probability sampling.

2 Estimation of the proportion of a sensitive attribute with a stratified unequal probability two stage randomized response model

In order to investigate the method of estimating the sensitive population proportion of the population, which is composed of N clusters with size M_i, we suggest a two-stage randomized response model with unequal probability sampling and with equal probability sampling by adapting Abdelfatah et al.'s procedure (2013).

We first investigate the estimation of the sensitive population proportion with unequal probability sampling with replacement (UPSWR) in Section 2.1, proportional to probability size without replacement (PPSWOR) in Section 2.2 and equal probability sampling in Section 2.3.

2.1 Estimation of a sensitive population proportion by UPSWR

Suppose the primary sampling units (PSUs) of size n clusters have been selected from the population of N clusters with size M_i with replacement, in which each ith PSU is selected with probability p_i and the secondary sampling units (SSUs) of size m_i $(i = 1, 2, \ldots, n)$ are selected from each chosen primary unit by SRSWR. Each respondent selected by the two-stage sampling procedure is requested to draw two cards simultaneously, one card from each deck of cards, and read the statements in order. Each interviewee in a SRSWR of m_i $(i = 1, 2, \ldots, n)$ respondents is provided with four decks of cards, as shown in Figure 1.

The respondent is requested to draw a card from Deck(3) only if directed by the outcome of Deck(1) and he/she is also requested to draw a card from Deck(4) only if directed by the outcome of Deck(2). Deck(3) and Deck(4) are exactly the same decks used by Abdelfatah et al. (2013). The respondent first matches his/her actual status with the question written on the card drawn from Deck(1) or Deck(3), and then he/she matches his/her actual status with the question written on the card

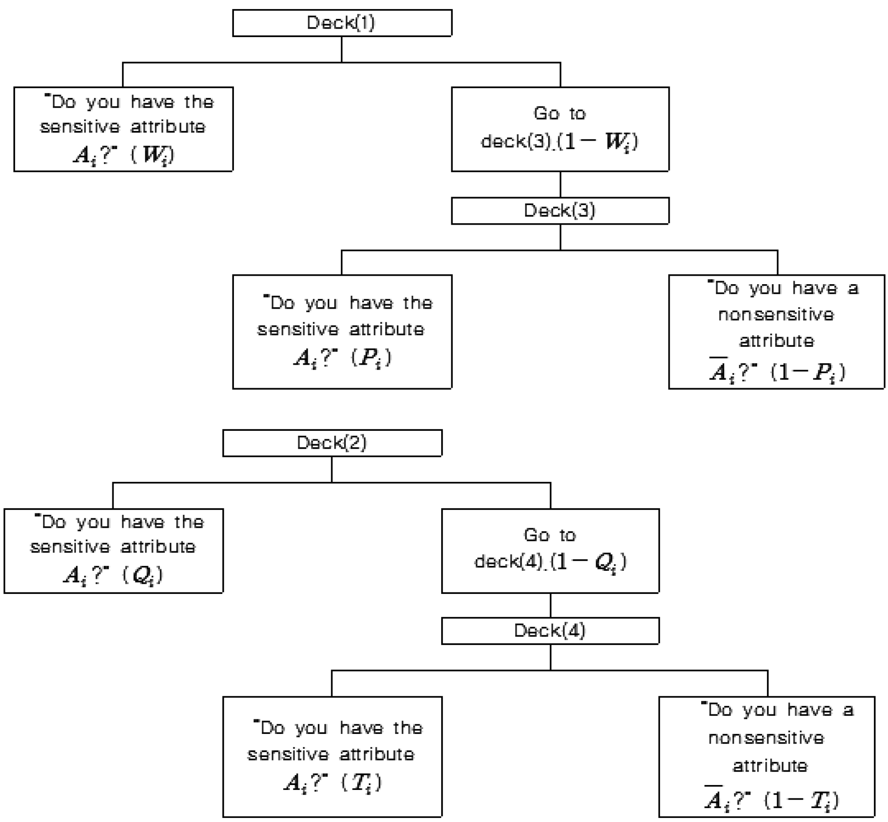

Figure 1 *The random device ith cluster.*

drawn from Deck(2) or Deck(4). The whole procedure is done completely by the respondent, away from the interviewer.

The probability of getting (Yes, Yes) response with sensitive population proportion π_i from ith cluster, θ_{i11}, is given by

$$\begin{aligned}\theta_{i11} &= P(Yes, Yes)\\ &= W_i Q_i \pi_i + W_i(1-Q_i)T_i\pi_i\\ &\quad + (1-W_i)P_i Q_i \pi_i + (1-W_i)P_i(1-Q_i)T_i\pi_i\\ &\quad + (1-W_i)(1-P_i)(1-Q_i)(1-T_i)(1-\pi_i)\\ &= \big[(1-W_i)P_i + (1-Q_i)T_i + Q_i + W_i - 1\big]\pi_i\\ &\quad + (1-W_i)(1-P_i)(1-Q_i)(1-T_i).\end{aligned}$$

In the same way, the probabilities, θ_{i10}, θ_{i01} and θ_{i00} are given by

$$\begin{aligned}\theta_{i10} &= P(Yes, No)\\ &= \big[W_i - Q_i + P_i(1-W_i) - T_i(1-Q_i)\big]\pi_i\\ &\quad + (1-W_i)(1-P_i)\big[Q_i + (1-Q_i)T_i\big],\\ \theta_{i01} &= P(No, Yes)\\ &= \big[Q_i - W_i + T_i(1-Q_i) - P_i(1-W_i)\big]\pi_i\\ &\quad + (1-Q_i)(1-T_i)\big[W_i + (1-W_i)P_i\big]\end{aligned}$$

and

$$\begin{aligned}\theta_{i00} &= P(No, No)\\ &= \big[1 - W_i - Q_i - P_i(1-W_i) - T_i(1-Q_i)\big]\pi_i\\ &\quad + W_i Q_i + W_i(1-Q_i)T_i + (1-W_i)P_i Q_i + (1-W_i)P_i(1-Q_i)T_i.\end{aligned}$$

The responses from the m_i respondents of ith cluster can be classified into a 2×2 contingency table as shown in Table 2.

In order to estimate the unknown population proportion π_i of the respondents belonging to sensitive group A_i in the ith cluster, let m_{i11}/m_i, m_{i10}/m_i, m_{i01}/m_i

Table 2 *Classification of the responses from the four decks of cards in ith cluster*

	Responses from Decks(2 or 4)		
Responses from Decks(1 or 3)	Yes	No	Σ
Yes	m_{i11}	m_{i10}	m_{i1+}
No	m_{i01}	m_{i00}	m_{i0+}
Σ	m_{i+1}	m_{i+0}	m_i

and m_{i00}/m_i be the observed proportions of (Yes, Yes), (Yes, No), (No, Yes) and (No, No) responses, and further let them be unbiased estimators for θ_{i11}, θ_{i10}, θ_{i01} and θ_{i00} respectively where $\sum_{j=0}^{1}\sum_{k=0}^{1}\theta_{ijk} = 1$.

We can define the squared distance between the observed proportions and the true proportions in each ith cluster as:

$$D_i = \frac{1}{2}\sum_{j=0}^{1}\sum_{k=0}^{1}\left(\theta_{ijk} - \frac{m_{ijk}}{m_i}\right)^2,$$

where

$$\begin{aligned}
D_i = \frac{1}{2}\Big[&\{(1-W_i)P_i + (1-Q_i)T_i + Q_i + W_i - 1\}\pi_i \\
&+ (1-W_i)(1-P_i)(1-Q_i)(1-T_i) - \frac{m_{i11}}{m_i}\Big]^2 \\
+ \frac{1}{2}\Big[&\{W_i - Q_i + P_i(1-W_i) - T_i(1-Q_i)\}\pi_i \\
&+ (1-W_i)(1-P_i)\{Q_i + (1-Q_i)T_i\} - \frac{m_{i10}}{m_i}\Big]^2 \\
+ \frac{1}{2}\Big[&\{Q_i - W_i + T_i(1-Q_i) - P_i(1-W_i)\}\pi_i \\
&+ (1-Q_i)(1-T_i)\{W_i + (1-W_i)P_i\} - \frac{m_{i01}}{m_i}\Big]^2 \\
+ \frac{1}{2}\Big[&\{1 - W_i - Q_i - P(1-W_i) - T_i(1-Q_i)\}\pi_i \\
&+ W_i Q_i + W_i(1-Q_i)T_i + (1-W_i)P_i Q_i \\
&+ (1-W_i)P_i(1-Q_i)T_i - \frac{m_{i00}}{m_i}\Big]^2.
\end{aligned}$$

To obtain π_i, which minimizes the squared distance D_i, we have

$$\begin{aligned}
\frac{\partial D_i}{\partial \pi_i} = {}& \left[(1-W_i)P_i + (1-Q_i)T_i + Q_i + W_i - 1\right]^2\pi_i \\
& - \frac{m_{i11}}{m_i}\left[(1-W_i)P_i + (1-Q_i)T_i + Q_i + W_i - 1\right] \\
& + \left[(1-W_i)P_i + (1-Q_i)T_i + Q_i + W_i - 1\right] \\
& \quad \times (1-W_i)(1-P_i)(1-Q_i)(1-T_i) \\
& + \left[W_i - Q_i + P_i(1-W_i) - T_i(1-Q_i)\right]^2\pi_i
\end{aligned}$$

$$
\begin{aligned}
&-\frac{m_{i10}}{m_i}\big[W_i - Q_i + P_i(1-W_i) - T_i(1-Q_i)\big]\\
&+\big[W_i - Q_i + P_i(1-W_i) - T_i(1-Q_i)\big]\\
&\quad\times(1-W_i)(1-P_i)\big[Q_i + (1-Q_i)T_i\big]\\
&+\big[Q_i - W_i + T_i(1-Q_i) - P_i(1-W_i)\big]^2\pi_i\\
&-\frac{m_{i01}}{m_i}\big[Q_i - W_i + T_i(1-Q_i) - P_i(1-W_i)\big]\\
&+\big[Q_i - W_i + T_i(1-Q_i) - P_i(1-W_i)\big]\\
&\quad\times(1-Q_i)(1-T_i)\big[W_i + (1-W_i)P_i\big]\\
&+\big[1 - W_i - Q_i - P_i(1-W_i) - T_i(1-Q_i)\big]^2\pi_i\\
&-\frac{m_{i00}}{m_i}\big[1 - W_i - Q_i - P_i(1-W_i) - T_i(1-Q_i)\big]\\
&+\big[1 - W_i - Q_i - P_i(1-W_i) - T_i(1-Q_i)\big]\\
&\quad\times\big[W_iQ_i + W_i(1-Q_i)T_i + (1-W_i)P_iQ_i + (1-W_i)P_i(1-Q_i)T_i\big]
\end{aligned}
$$

and setting $\frac{\partial D_i}{\partial \pi_i} = 0$, we obtain the following estimator $\hat{\pi}_i$ of the population proportion π_i in ith cluster

$$\hat{\pi}_i = \frac{1}{2} + \frac{(m_{i11}/m_i - m_{i00}/m_i)B_i + (m_{i10}/m_i - m_{i01}/m_i)C_i}{2(B_i^2 + C_i^2)}, \tag{2.1}$$

where $B_i = (1-W_i)P_i + (1-Q_i)T_i + W_i + Q_i - 1$, $C_i = W_i - Q_i + (1-W_i)P_i - (1-Q_i)T_i$.

Thus, the overall estimator $\hat{\pi}_{\text{upswr}}$ of the population proportion π is obtained by

$$
\begin{aligned}
\hat{\pi}_{\text{upswr}} &= \frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i\hat{\pi}_i}{p_i}\\
&= \frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i}{p_i} \qquad (2.2)\\
&\qquad\times\left[\frac{1}{2} + \frac{(m_{i11}/m_i - m_{i00}/m_i)B_i + (m_{i10}/m_i - m_{i01}/m_i)C_i}{2(B_i^2 + C_i^2)}\right],
\end{aligned}
$$

where $M_0 = \sum_{i=1}^{N} M_i$.

Theorem 2.1. *The $\hat{\pi}_{\text{upswr}}$ is an unbiased estimator of the sensitive population proportion π.*

Proof. The expected value of $\hat{\pi}_{\text{upswr}}$ is given by

$$E_1E_2(\hat{\pi}_{\text{upswr}}) = E_1E_2\left[\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i\hat{\pi}_i}{p_i}\right]$$
$$= E_1\left[\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i}{p_i}E_2(\hat{\pi}_i)\right].$$

It follows from the fact that $E(m_{ijk}/m) = \theta_{ijk}$, $i = 1, 2, \ldots, n$; $j = 0, 1$; $k = 0, 1$.

$$E_2(\hat{\pi}_i) = \left[\frac{1}{2} + \frac{(m_{i11}/m_i - m_{i00}/m_i)B_i + (m_{i10}/m_i - m_{i01}/m_i)C_i}{2(B_i^2 + C_i^2)}\right]$$
$$= \frac{1}{2} + \frac{(\theta_{i11} - \theta_{i00})B_i + (\theta_{i10} - \theta_{i01})C_i}{2(B_i^2 + C_i^2)}$$
$$= \pi_i.$$

Hence, we can prove Theorem 2.1

$$E_1E_2(\hat{\pi}_{\text{upswr}}) = E_1\left[\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i\pi_i}{p_i}\right]$$
$$= \frac{1}{M_0}\sum_{i=1}^{N}p_i\frac{M_i\pi_i}{p_i}$$
$$= \pi. \qquad \square$$

Theorem 2.2. *The variance of the estimator $\hat{\pi}_{\text{upswr}}$ is given by*

$$V(\hat{\pi}_{\text{upswr}}) = \frac{1}{nM_0^2}\sum_{i=1}^{N}p_i\left[\frac{M_i\pi_i}{p_i} - M_0\pi\right]^2$$
$$+ \frac{1}{nM_0^2}\sum_{i=1}^{N}\frac{M_i^2}{p_i}\frac{1}{4m_i} \tag{2.3}$$
$$\times\left[\frac{B_i^2(E_i + F_i) + C_i^2(G_i + H_i)}{(B_i^2 + C_i^2)^2} - (2\pi_i - 1)^2\right],$$

where

$$E_i = W_iQ_i + W_i(1 - Q_i)T_i + (1 - W_i)P_iQ_i + (1 - W_i)P_i(1 - Q_i)T_i,$$
$$F_i = (1 - W_i)(1 - P_i)(1 - Q_i)(1 - T_i),$$
$$G_i = (1 - Q_i)(1 - T_i)[W_i + (1 - W_i)P_i],$$
$$H_i = (1 - W_i)(1 - P_i)[Q_i + (1 - Q_i)T_i].$$

Proof.

$$V(\hat{\pi}_{\text{upswr}}) = V_1 E_2(\hat{\pi}_{\text{upswr}}) + E_1 V_2(\hat{\pi}_{\text{upswr}}).$$

We can see

$$\begin{aligned} V_1 E_2(\hat{\pi}_{\text{upswr}}) &= V_1 E_2\left[\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i\hat{\pi}_i}{p_i}\right] \\ &= V_1\left[\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i\pi_i}{p_i}\right] \\ &= \frac{1}{nM_0^2}\sum_{i=1}^{N} p_i\left[\frac{M_i\pi_i}{p_i} - M_0\pi\right]^2 \end{aligned}$$

and

$$\begin{aligned} &E_1 V_2(\hat{\pi}_{\text{upswr}}) \\ &\quad = E_1 V_2\left[\frac{1}{nM_0}\sum_{i=1}^{n}\frac{M_i\hat{\pi}_i}{p_i}\right] \\ &\quad = E_1\left[\frac{1}{(nM_0)^2}\sum_{i=1}^{n}\frac{M_i^2}{p_i^2}V_2(\hat{\pi}_i)\right] \\ &\quad = E_1\left[\frac{1}{(nM_0)^2}\right. \\ &\qquad\quad \times \sum_{i=1}^{n}\frac{M_i^2}{p_i^2}V_2\left\{\frac{1}{2} + \big((m_{i11}/m_i - m_{i00}/m_i)B_i\right. \\ &\qquad\qquad\qquad \left.\left. + (m_{i10}/m_i - m_{i01}/m_i)C_i\big)/\big(2\big(B_i^2 + C_i^2\big)\big)\right\}\right] \\ &\quad = E_1\left[\frac{1}{(nM_0)^2}\sum_{i=1}^{n}\frac{M_i^2}{p_i^2}\frac{1}{4(B_i^2 + C_i^2)^2}\right. \\ &\qquad\qquad \times\left\{B_i^2 \times V\left(\frac{m_{i11}}{m_i} - \frac{m_{i00}}{m_i}\right) + C_i^2 \times V\left(\frac{m_{i10}}{m_i} - \frac{m_{i01}}{m_i}\right)\right. \\ &\qquad\qquad\quad \left.\left. + 2B_iC_i \times \operatorname{Cov}\left(\frac{m_{i11} - m_{i00}}{m_i}, \frac{m_{i10} - m_{i01}}{m_i}\right)\right\}\right]. \end{aligned}$$

Using the following results from the multinomial distribution,

$$V(m_{i11}/m_i) = \theta_{i11}(1 - \theta_{i11})/m_i;$$

$$\operatorname{Cov}(m_{i11}/m_i, m_{i10}/m_i) = -\theta_{i11}\theta_{i10}/m_i;$$

$$V(m_{i10}/m_i) = \theta_{i10}(1-\theta_{i10})/m_i;$$
$$\mathrm{Cov}(m_{i10}/m_i, m_{i01}/m_i) = -\theta_{i10}\theta_{i01}/m_i;$$
$$V(m_{i01}/m_i) = \theta_{i01}(1-\theta_{i01})/m_i;$$
$$\mathrm{Cov}(m_{i11}/m_i, m_{i01}/m_i) = -\theta_{i11}\theta_{i01}/m_i;$$
$$V(m_{i00}/m_i) = \theta_{i00}(1-\theta_{i00})/m_i;$$
$$\mathrm{Cov}(m_{i10}/m_i, m_{i00}/m_i) = -\theta_{i10}\theta_{i00}/m_i;$$
$$\mathrm{Cov}(m_{i01}/m_i, m_{i00}/m_i) = -\theta_{i01}\theta_{i00}/m_i;$$
$$\mathrm{Cov}(m_{i11}/m_i, m_{i00}/m_i) = -\theta_{i11}\theta_{i00}/m_i.$$

We can rewrite $E_1 V_2(\hat{\pi}_{\mathrm{upswr}})$ as follows and prove Theorem 2.2.

$$\begin{aligned}
E_1 V_2(\hat{\pi}_{\mathrm{upswr}}) &\\
&= E\left[\frac{1}{(nM_0)^2}\sum_{i=1}^{n}\frac{M_i^2}{p_i^2}\left\{\frac{B_i^2(E_i+F_i)+C_i^2(G_i+H_i)}{4m_i(B_i^2+C_i^2)^2} - \frac{(2\pi_i-1)^2}{4m_i}\right\}\right]\\
&= \frac{n}{(nM_0)^2}\sum_{i=1}^{N} p_i \frac{M_i^2}{p_i^2}\left[\frac{B_i^2(E_i+F_i)+C_i^2(G_i+H_i)}{4m_i(B_i^2+C_i^2)^2} - \frac{(2\pi_i-1)^2}{4m_i}\right]\\
&= \frac{1}{nM_0^2}\sum_{i=1}^{N}\frac{M_i^2}{p_i}\frac{1}{4m_i}\left[\frac{B_i^2(E_i+F_i)+C_i^2(G_i+H_i)}{(B_i^2+C_i^2)^2} - (2\pi_i-1)^2\right]. \qquad \square
\end{aligned}$$

Also, an unbiased estimator of the variance of $\hat{\pi}_{\mathrm{upswr}}$ is given by

$$\begin{aligned}
\hat{V}(\hat{\pi}_{\mathrm{upswr}}) &= \frac{1}{nM_0^2}\sum_{i=1}^{n} p_i\left[\frac{M_i\hat{\pi}_i}{p_i} - M_0\hat{\pi}_{\mathrm{upswr}}\right]^2\\
&\quad + \frac{1}{nM_0^2}\sum_{i=1}^{n}\frac{M_i^2}{p_i}\frac{1}{4(m_i-1)} \qquad (2.4)\\
&\qquad \times\left[\frac{B_i^2(E_i+F_i)+C_i^2(G_i+H_i)}{(B_i^2+C_i^2)^2} - (2\hat{\pi}_i-1)^2\right].
\end{aligned}$$

Meanwhile, if the M_i is known and each unit of n PSUs is selected with probability proportional to its size M_i, then the unequal probability $p_i = M_i/M_0$. We call it sampling with probability proportional to size, or *pps* sampling.

In this case, the estimator $\hat{\pi}_{\mathrm{ppswr}}$ of the population proportion π is given by

$$\begin{aligned}
\hat{\pi}_{\mathrm{ppswr}} &= \frac{1}{n}\sum_{i=1}^{n}\hat{\pi}_i\\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{2} + \frac{(m_{i11}/m_i - m_{i00}/m_i)B_i + (m_{i10}/m_i - m_{i01}/m_i)C_i}{2(B_i^2+C_i^2)}\right].
\end{aligned} \qquad (2.5)$$

The variance and variance estimator of $\hat{\pi}_{\text{ppswr}}$ are given respectively by

$$
\begin{aligned}
V(\hat{\pi}_{\text{ppswr}}) = {} & \frac{1}{nM_0}\sum_{i=1}^{N} M_i(\pi_i-\pi)^2 \\
& + \frac{1}{nM_0}\sum_{i=1}^{N} M_i \frac{1}{4m_i} \\
& \times \left[\frac{B_i^2(E_i+F_i)+C_i^2(G_i+H_i)}{(B_i^2+C_i^2)^2} - (2\pi_i-1)^2\right],
\end{aligned}
\tag{2.6}
$$

$$
\begin{aligned}
\hat{V}(\hat{\pi}_{\text{ppswr}}) = {} & \frac{1}{nM_0}\sum_{i=1}^{n} M_i(\hat{\pi}_i-\hat{\pi}_{\text{ppswr}})^2 \\
& + \frac{1}{nM_0}\sum_{i=1}^{n} M_i \frac{1}{4(m_i-1)} \\
& \times \left[\frac{B_i^2(E_i+F_i)+C_i^2(G_i+H_i)}{(B_i^2+C_i^2)^2} - (2\hat{\pi}_i-1)^2\right].
\end{aligned}
\tag{2.7}
$$

2.2 Estimation of a sensitive population proportion by PPSWOR

Suppose the primary sampling units (PSUs) of size n have been selected from the population of N clusters with size M_i by proportional to probability size without replacement (PPSWOR), and the secondary sampling units (SSUs) of size m_i ($i = 1, 2, \dots, n$) are selected from each chosen primary unit by simple random sampling with replacement (SRSWR).

The estimator $\hat{\pi}_{\text{ppswor}}$ of sensitive population proportion π can be obtained by applying Abdelfatah et al. (2013) model

$$
\hat{\pi}_{\text{ppswor}} = \frac{1}{M_0}\sum_{i=1}^{n}\frac{M_i\hat{\pi}_i}{\theta_i},
\tag{2.8}
$$

where θ_i is the probability that unit i is included in the sample.

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

$$
\begin{aligned}
V(\hat{\pi}_{\text{ppswor}}) = {} & \frac{1}{M_0^2}\sum_{i=1}^{N}\sum_{j>i}^{N}(\theta_i\theta_j-\theta_{ij})\left[\frac{M_i\pi_i}{\theta_i}-\frac{M_j\pi_j}{\theta_j}\right]^2 \\
& + \frac{1}{M_0^2}\sum_{i=1}^{N}\frac{M_i^2}{\theta_i}\frac{1}{4m_i} \\
& \times \left[\frac{B_i^2(E_i+F_i)+C_i^2(G_i+H_i)}{(B_i^2+C_i^2)^2} - (2\pi_i-1)^2\right],
\end{aligned}
\tag{2.9}
$$

where θ_{ij} is the probability that unit i and j are included in the sample.

The estimator of $V(\hat{\pi}_{\text{ppswor}})$ is given by

$$\hat{V}(\hat{\pi}_{\text{ppswor}}) = \frac{1}{M_0^2}\sum_{i=1}^{n}\sum_{j>i}^{n}\frac{(\theta_i\theta_j - \theta_{ij})}{\theta_{ij}}\left[\frac{M_i\hat{\pi}_i}{\theta_i} - \frac{M_j\hat{\pi}_j}{\theta_j}\right]^2 + \frac{1}{M_0^2}\sum_{i=1}^{N}\frac{M_i^2}{\theta_i}\frac{1}{4(m_i-1)} \times \left[\frac{B_i^2(E_i+F_i)+C_i^2(G_i+H_i)}{(B_i^2+C_i^2)^2} - (2\hat{\pi}_i - 1)^2\right]. \tag{2.10}$$

2.3 Estimation of a sensitive population proportion by equal two-stage sampling

Suppose the primary sampling units (PSUs) of size n have been selected from the population of N clusters with size M_i by SRSWR and the secondary sampling units (SSUs) of size m_i $(i = 1, 2, \dots, n)$ are selected from each chosen primary unit by simple random sampling with replacement (SRSWR).

The estimator $\hat{\pi}_{\text{wr}}$ of sensitive population proportion π can be obtained by applying Abdelfatah et al. (2013) model

$$\hat{\pi}_{\text{wr}} = \frac{N}{M_0 n}\sum_{i=1}^{n} M_i\hat{\pi}_i, \tag{2.11}$$

where $\hat{\pi}_i$ is the estimator of the population proportion π_i in ith cluster as in (2.1).

The variance of $\hat{\pi}_{\text{wr}}$ and its estimator are given as follows.

$$V(\hat{\pi}_{\text{wr}}) = N^2\frac{1}{nM_0^2}\frac{1}{N-1}\sum_{i=1}^{N}(M_i\pi_i - \bar{M}\pi)^2 + \frac{N}{nM_0^2}\sum_{i=1}^{N}M_i^2\frac{1}{4m_i}\left[\frac{B_i^2(E_i+F_i)+C_i^2(G_i+H_i)}{(B_i^2+C_i^2)^2} - (2\pi_i - 1)^2\right] \tag{2.12}$$

and

$$\hat{V}(\hat{\pi}_{\text{wr}}) = N^2\frac{1}{nM_0^2}\frac{1}{n-1}\sum_{i=1}^{n}(M_i\hat{\pi}_i - \bar{M}\hat{\pi}_{\text{wr}})^2 + \frac{N}{nM_0^2}\sum_{i=1}^{n}M_i^2\frac{1}{4(m_i-1)} \times \left[\frac{B_i^2(E_i+F_i)+C_i^2(G_i+H_i)}{(B_i^2+C_i^2)^2} - (2\hat{\pi}_i - 1)^2\right], \tag{2.13}$$

where $\bar{M} = \frac{M_0}{N}$.

2.4 A comparison of PPSWR sampling and equal two-stage sampling

When we set $N-1 \fallingdotseq N$, the difference between the variance of (2.12) and (2.6) can be approximated as follows.

$$\begin{aligned} V(\hat{\pi}_{\text{wr}}) &- V(\hat{\pi}_{\text{ppswr}}) \\ &= \frac{1}{nM_0\bar{M}}\Bigg[\sum_{i=1}^{N}(M_i-\bar{M})^2\pi_i^2 + \bar{M}\sum_{i=1}^{N}(M_i-\bar{M})\pi_i^2 \\ &\quad + \sum_{i=1}^{N}(M_i-\bar{M})^2\frac{1}{4m_i} \\ &\quad \times\left(\frac{B_i^2(E_i+F_i)+C_i^2(G_i+H_i)}{(B_i^2+C_i^2)^2}-(2\pi_i-1)^2\right) \\ &\quad + \bar{M}\sum_{i=1}^{N}(M_i-\bar{M})\frac{1}{4m_i} \\ &\quad \times\left(\frac{B_i^2(E_i+F_i)+C_i^2(G_i+H_i)}{(B_i^2+C_i^2)^2}-(2\pi_i-1)^2\right)\Bigg]. \end{aligned} \tag{2.14}$$

If $M_i=\bar{M}=\frac{M_0}{N}$ in (2.14), then $V(\hat{\pi}_{\text{wr}})=V(\hat{\pi}_{\text{ppswr}})$. That is, if the sizes of the clusters are equal, then the selection probabilities are equal to $\frac{1}{N}$, and they are equal to the selection probabilities of equal two-stage with replacement sampling. Hence, the efficiency of the two methods is equivalent.

If the cluster sizes M_i are different significantly on the right of (2.14), the first term $\sum_{i=1}^{N}(M_i-\bar{M})^2\pi_i^2$ and the third term have value greater than zero and the second and fourth term have relatively small value. Therefore, the estimation by ppswr sampling is more profitable than that of equal two-stage with replacement sampling when the cluster sizes are unequal.

3 A stratified *pps* estimation of a sensitive attribute by two randomized response model

When the population is composed of a number of different-sized strata, we deal with the estimation of the sensitive attribute of the population by applying stratified unequal probability sampling (SUPS) or equal probability sampling to Abdelfatah et al. (2013) model.

3.1 Estimation of sensitive population proportion by SUPSWR

Let the population be compose of a number of mutually disjoint L strata of N_h $(h=1,2,\dots,L)$, where each stratum is consisted of N_h clusters of size M_{hi}. The

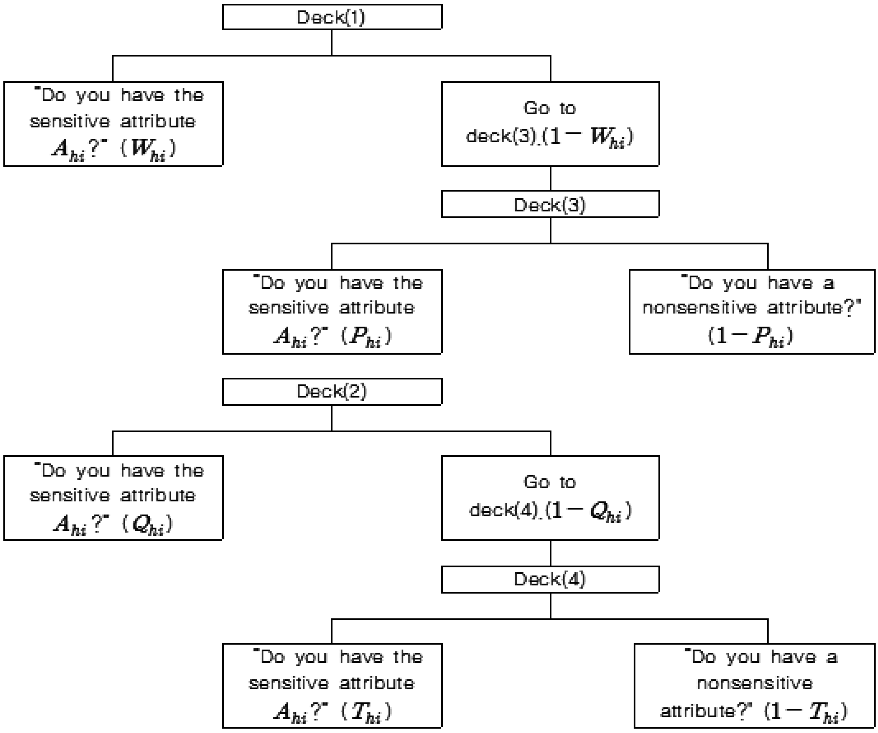

Figure 2 *The random device of stratum* i*th cluster of stratum* h*.*

n_h clusters are selected with replacement by unequal probabilities p_{hi} from the hth stratum, in which m_{hi} $(i = 1, 2, \ldots, n_h)$ observation units are selected by SRSWR from each cluster.

Each of m_{hi} respondents selected by SRSWR from the ith cluster of size M_{hi} are requested to draw two cards simultaneously, one card from each deck of cards, and read the statements in order. The respondent is requested to draw a card from Deck(3) only if directed by the outcome of Deck(1), and he/she is also requested to draw a card from Deck(4) only if directed by the outcome of Deck(2). Deck(3) and Deck(4) are exactly the same decks used by Abdelfatah et al. (2013). The respondent first matches his/her actual status with the statement (question) written on the card drawn from Deck(1) or Deck(3), and then he/she matches his/her actual status with the statement (question) written on the card drawn from Deck(2) or Deck(4), as shown in Figure 2. The whole procedure is done completely by the respondent, away from the interviewer.

Since the response (Yes, Yes) from ith cluster in stratum h can be answered from any respondent regardless of having sensitive attribute A_{hi}, the interviewer can't know the interviewee's real status.

The probability of getting (Yes, Yes) response from ith cluster in stratum h, θ_{hi11}, is given by

$$\begin{aligned}\theta_{hi11} &= P(Yes, Yes)\\ &= W_{hi}Q_{hi}\pi_{hi} + W_{hi}(1-Q_{hi})T_{hi}\pi_{hi} + (1-W_{hi})P_{hi}Q_{hi}\pi_{hi}\\ &\quad + (1-W_{hi})P_{hi}(1-Q_{hi})T_{hi}\pi_{hi}\\ &\quad + (1-W_{hi})(1-P_{hi})(1-Q_{hi})(1-T_{hi})(1-\pi_{hi})\\ &= \big[(1-W_{hi})P_{hi} + (1-Q_{hi})T_{hi} + Q_{hi} + W_{hi} - 1\big]\pi_{hi}\\ &\quad + (1-W_{hi})(1-P_{hi})(1-Q_{hi})(1-T_{hi}).\end{aligned}$$

In the same way, the probabilities, θ_{hi10}, θ_{hi01}, and θ_{hi00} are given by

$$\begin{aligned}\theta_{hi10} &= P(Yes, No)\\ &= \big[W_{hi} - Q_{hi} + P_{hi}(1-W_{hi}) - T_{hi}(1-Q_{hi})\big]\pi_{hi}\\ &\quad + (1-W_{hi})(1-P_{hi})\big[Q_{hi} + (1-Q_{hi})T_{hi}\big],\\ \theta_{hi01} &= P(No, Yes)\\ &= \big[Q_{hi} - W_{hi} + T_{hi}(1-Q_{hi}) - P_{hi}(1-W_{hi})\big]\pi_{hi}\\ &\quad + (1-Q_{hi})(1-T_{hi})\big[W_{hi} + (1-W_{hi})P_{hi}\big]\end{aligned}$$

and

$$\begin{aligned}\theta_{hi00} &= P(No, No)\\ &= \big[1 - W_{hi} - Q_{hi} - P_{hi}(1-W_{hi}) - T_{hi}(1-Q_{hi})\big]\pi_{hi}\\ &\quad + W_{hi}Q_{hi} + W_{hi}(1-Q_{hi})T_{hi} + (1-W_{hi})P_{hi}Q_{hi}\\ &\quad + (1-W_{hi})P_{hi}(1-Q_{hi})T_{hi}.\end{aligned}$$

The responses from the m_{hi} respondents from the ith cluster in stratum h can be classified into a 2×2 contingency table as shown in Table 3. In order to estimate the unknown population proportion π_{hi} of the respondents belonging to the sensitive group A_{hi} in ith cluster of stratum h, let m_{hi11}/m_{hi}, m_{hi10}/m_{hi}, m_{hi01}/m_{hi}

Table 3 *Classification of the responses from the four decks of cards in ith cluster of stratum h*

	Responses from Decks(2 or 4)		
Responses from Decks(1 or 3)	Yes	No	Σ
Yes	m_{hi11}	m_{hi10}	m_{hi1+}
No	m_{hi01}	m_{hi00}	m_{hi0+}
Σ	m_{hi+1}	m_{hi+0}	m_{hi}

and m_{hi00}/m_{hi} be the observed proportions of (Yes, Yes), (Yes, No), (No, Yes) and (No, No) responses and, and further let them be unbiased estimators for θ_{hi11}, θ_{hi10}, θ_{hi01} and θ_{hi00}, respectively where $\sum_{j=0}^{1}\sum_{k=0}^{1}\theta_{hijk}=1$.

We can define the squared distance between the observed proportions and the true proportions in the ith cluster of stratum h as:

$$D_{hi}=\frac{1}{2}\sum_{i=1}^{n_h}\sum_{j=0}^{1}\sum_{k=0}^{1}\left(\theta_{hijk}-\frac{m_{hijk}}{m_{hi}}\right)^2,$$

where

$$\begin{aligned}
D_{hi}=\frac{1}{2}\Big[&\big\{(1-W_{hi})P_{hi}+(1-Q_{hi})T_{hi}+Q_{hi}+W_{hi}-1\big\}\pi_{hi}\\
&+(1-W_{hi})(1-P_{hi})(1-Q_{hi})(1-T_{hi})-\frac{m_{hi11}}{m_{hi}}\Big]^2\\
&+\frac{1}{2}\Big[\big\{W_{hi}-Q_{hi}+P_{hi}(1-W_{hi})-T_{hi}(1-Q_{hi})\big\}\pi_{hi}\\
&\quad+(1-W_{hi})(1-P_{hi})\big\{Q_{hi}+(1-Q_{hi})T_{hi}\big\}-\frac{m_{hi10}}{m_{hi}}\Big]^2\\
&+\frac{1}{2}\Big[\big\{Q_{hi}-W_{hi}+T_{hi}(1-Q_{hi})-P_{hi}(1-W_{hi})\big\}\pi_{hi}\\
&\quad+(1-Q_{hi})(1-T_{hi})\big\{W_{hi}+(1-W_{hi})P_{hi}\big\}-\frac{m_{hi01}}{m_{hi}}\Big]^2\\
&+\frac{1}{2}\Big[\big\{1-W_{hi}-Q_{hi}-P(1-W_{hi})-T_{hi}(1-Q_{hi})\big\}\pi_{hi}\\
&\quad+W_{hi}Q_{hi}+W_{hi}(1-Q_{hi})T_{hi}+(1-W_{hi})P_{hi}Q_{hi}\\
&\quad+(1-W_{hi})P_{hi}(1-Q_{hi})T_{hi}-\frac{m_{hi00}}{m_{hi}}\Big]^2.
\end{aligned}$$

To obtain π_{hi} that minimizes the squared distance D_{hi}, we have

$$\begin{aligned}
\frac{\partial D_{hi}}{\partial \pi_{hi}}=&\big[(1-W_{hi})P_{hi}+(1-Q_{hi})T_{hi}+Q_{hi}+W_{hi}-1\big]^2\pi_{hi}\\
&-\frac{m_{hi11}}{m_{hi}}\big[(1-W_{hi})P_{hi}+(1-Q_{hi})T_{hi}+Q_{hi}+W_{hi}-1\big]\\
&+\big[(1-W_{hi})P_{hi}+(1-Q_{hi})T_{hi}+Q_{hi}+W_{hi}-1\big]\\
&\quad\times(1-W_{hi})(1-P_{hi})(1-Q_{hi})(1-T_{hi})\\
&+\big[W_{hi}-Q_{hi}+P_{hi}(1-W_{hi})-T_{hi}(1-Q_{hi})\big]^2\pi_{hi}
\end{aligned}$$

$$
\begin{aligned}
&- \frac{m_{hi10}}{m_{hi}}\big[W_{hi} - Q_{hi} + P_{hi}(1 - W_{hi}) - T_{hi}(1 - Q_{hi})\big] \\
&+ \big[W_{hi} - Q_{hi} + P_{hi}(1 - W_{hi}) - T_{hi}(1 - Q_{hi})\big] \\
&\quad \times (1 - W_{hi})(1 - P_{hi})\big[Q_{hi} + (1 - Q_{hi})T_{hi}\big] \\
&+ \big[Q_{hi} - W_{hi} + T_{hi}(1 - Q_{hi}) - P_{hi}(1 - W_{hi})\big]^2 \pi_{hi} \\
&- \frac{m_{hi01}}{m_{hi}}\big[Q_{hi} - W_{hi} + T_{hi}(1 - Q_{hi}) - P_{hi}(1 - W_{hi})\big] \\
&+ \big[Q_{hi} - W_{hi} + T_{hi}(1 - Q_{hi}) - P_{hi}(1 - W_{hi})\big] \\
&\quad \times (1 - Q_{hi})(1 - T_{hi})\big[W_{hi} + (1 - W_{hi})P_{hi}\big] \\
&+ \big[1 - W_{hi} - Q_{hi} - P_{hi}(1 - W_{hi}) - T_{hi}(1 - Q_{hi})\big]^2 \pi_{hi} \\
&- \frac{m_{hi00}}{m_{hi}}\big[1 - W_{hi} - Q_{hi} - P_{hi}(1 - W_{hi}) - T_{hi}(1 - Q_{hi})\big] \\
&+ \big[1 - W_{hi} - Q_{hi} - P_{hi}(1 - W_{hi}) - T_{hi}(1 - Q_{hi})\big] \\
&\quad \times \big[W_{hi}Q_{hi} + W_{hi}(1 - Q_{hi})T_{hi} \\
&\qquad + (1 - W_{hi})P_{hi}Q_{hi} + (1 - W_{hi})P_{hi}(1 - Q_{hi})T_{hi}\big],
\end{aligned}
$$

and setting $\frac{\partial D_{hi}}{\partial \pi_{hi}} = 0$, we obtain the following estimator $\hat{\pi}_{hi}$ of the population proportion π_{hi} in the ith cluster of stratum h

$$
\hat{\pi}_{hi} = \frac{1}{2} + \frac{(m_{hi11}/m_{hi} - m_{hi00}/m_{hi})B_{hi} + (m_{hi10}/m_{hi} - m_{hi01}/m_{hi})C_{hi}}{2(B_{hi}^2 + C_{hi}^2)}, \quad (3.1)
$$

where $B_{hi} = (1 - W_{hi})P_{hi} + (1 - Q_{hi})T_{hi} + W_{hi} + Q_{hi} - 1$, $C_{hi} = W_{hi} - Q_{hi} + (1 - W_{hi})P_{hi} - (1 - Q_{hi})T_{hi}$.

Thus, the estimator $\hat{\pi}_h$ of the population proportion of hth stratum π_h is obtained by

$$
\begin{aligned}
\hat{\pi}_h &= \frac{1}{n_h M_{h0}} \sum_{i=1}^{n_h} \frac{M_{hi}\hat{\pi}_{hi}}{p_{hi}} \\
&= \frac{1}{n_h M_{h0}} \sum_{i=1}^{n_h} \frac{M_{hi}}{p_{hi}} \qquad (3.2) \\
&\quad \times \left[\frac{1}{2} + \frac{(m_{hi11}/m_{hi} - m_{hi00}/m_{hi})B_{hi} + (m_{hi10}/m_{hi} - m_{hi01}/m_{hi})C_{hi}}{2(B_{hi}^2 + C_{hi}^2)}\right],
\end{aligned}
$$

where $M_{h0} = \sum_{i=1}^{N_h} M_{hi}$.

Therefore, the overall estimator $\hat{\pi}_{\text{supswr}}$ of the population proportion π is given by

$$\begin{aligned}\hat{\pi}_{\text{supswr}} &= \sum_{h=1}^{L} Z_h \frac{1}{n_h M_{h0}} \\ &\times \sum_{i=1}^{n_h} \frac{M_{hi}}{p_{hi}} \\ &\times \left[\frac{1}{2} + \frac{(m_{hi11}/m_{hi} - m_{hi00}/m_{hi})B_{hi} + (m_{hi10}/m_{hi} - m_{hi01}/m_{hi})C_{hi}}{2(B_{hi}^2 + C_{hi}^2)}\right],\end{aligned} \tag{3.3}$$

where $Z_h = \frac{N_h}{N}$.

Theorem 3.1. *The estimator $\hat{\pi}_{\text{supswr}}$ is an unbiased estimator of the population proportion π.*

Proof.

$$\begin{aligned}E_1 E_2(\hat{\pi}_{\text{supswr}}) &= E_1 E_2\left[\sum_{h=1}^{L} Z_h \frac{1}{n_h M_{h0}} \sum_{i=1}^{n_h} \frac{M_{hi}\hat{\pi}_{hi}}{p_{hi}}\right] \\ &= E_1\left[\sum_{h=1}^{L} Z_h \frac{1}{n_h M_{h0}} \sum_{i=1}^{n_h} \frac{M_{hi}}{p_{hi}} E_2(\hat{\pi}_{hi})\right]\end{aligned}$$

and $E(m_{hijk}/m_h) = \theta_{hijk}$, $h = 1, 2, \ldots, L$; $i = 1, 2, \ldots, n_h$; $j = 0, 1$; $k = 0, 1$.

$$\begin{aligned}E_2(\hat{\pi}_{hi}) &= \left[\frac{1}{2} + \frac{(m_{hi11}/m_{hi} - m_{hi00}/m_{hi})B_{hi} + (m_{hi10}/m_{hi} - m_{hi01}/m_{hi})C_{hi}}{2(B_{hi}^2 + C_{hi}^2)}\right] \\ &= \frac{1}{2} + \frac{(\theta_{hi11} - \theta_{hi00})B_{hi} + (\theta_{hi10} - \theta_{hi01})C_{hi}}{2(B_{hi}^2 + C_{hi}^2)} = \pi_{hi}.\end{aligned}$$

Hence, we can show

$$\begin{aligned}E_1 E_2(\hat{\pi}_{\text{supswr}}) &= E_1\left[\sum_{h=1}^{L} Z_h \frac{1}{n_h M_{h0}} \sum_{i=1}^{n_h} \frac{M_{hi}\pi_{hi}}{p_{hi}}\right] \\ &= \sum_{h=1}^{L} Z_h \frac{1}{M_{h0}} \sum_{i=1}^{N_h} p_{hi} \frac{M_{hi}\pi_{hi}}{p_{hi}} = \sum_{h=1}^{L} Z_h \pi_h = \pi.\end{aligned}$$

□

Theorem 3.2. *When n_h clusters are selected with p_{hi} from hth stratum of size M_{hi}, in which m_h ($i = 1, 2, \ldots, n_h$) observation units are selected by SRSWR from* *each cluster. The variance of the estimator $\hat{\pi}_{\text{supswr}}$ is given by*

$$V(\hat{\pi}_{\text{supswr}}) = \sum_{h=1}^{L} Z_h^2 \Bigg[\frac{1}{n_h M_{h0}^2} \sum_{i=1}^{N_h} p_{hi} \left(\frac{M_{hi}\pi_{hi}}{p_{hi}} - M_{h0}\pi_h \right)^2 + \frac{1}{n_h M_{h0}^2} \sum_{i=1}^{N_h} \frac{M_{hi}^2}{p_{hi}} \frac{1}{4m_{hi}} \times \left\{ \frac{B_{hi}^2 (E_{hi} + F_{hi}) + C_{hi}^2 (G_{hi} + H_{hi})}{(B_{hi}^2 + C_{hi}^2)^2} - (2\pi_{hi} - 1)^2 \right\} \Bigg], \tag{3.4}$$

where

$$\begin{aligned} E_{hi} &= W_{hi} Q_{hi} + W_{hi}(1 - Q_{hi}) T_{hi} + (1 - W_{hi}) P_{hi} Q_{hi} \\ &\quad + (1 - W_{hi}) P_{hi} (1 - Q_{hi}) T_{hi}, \\ F_{hi} &= (1 - W_{hi})(1 - P_{hi})(1 - Q_{hi})(1 - T_{hi}), \\ G_{hi} &= (1 - Q_{hi})(1 - T_{hi}) \big[W_{hi} + (1 - W_{hi}) P_{hi} \big], \\ H_{hi} &= (1 - W_{hi})(1 - P_{hi}) \big[Q_{hi} + (1 - Q_{hi}) T_{hi} \big]. \end{aligned}$$

Proof.

$$V(\hat{\pi}_{\text{supswr}}) = V_1 E_2(\hat{\pi}_{\text{supswr}}) + E_1 V_2(\hat{\pi}_{\text{supswr}}),$$

where

$$\begin{aligned} V_1 E_2(\hat{\pi}_{\text{supswr}}) &= V_1 E_2 \left[\sum_{h=1}^{L} Z_h \frac{1}{n_h M_{h0}} \sum_{i=1}^{n_h} \frac{M_{hi} \hat{\pi}_{hi}}{p_{hi}} \right] \\ &= V_1 \left[\sum_{h=1}^{L} Z_h \frac{1}{n_h M_{h0}} \sum_{i=1}^{n_h} \frac{M_{hi} \pi_{hi}}{p_{hi}} \right] \\ &= \sum_{h=1}^{L} Z_h^2 \frac{1}{n_h M_0^2} \sum_{i=1}^{N_h} p_{hi} \left[\frac{M_{hi}\pi_{hi}}{p_{hi}} - M_{h0}\pi_h \right]^2 \end{aligned}$$

and

$$E_1 V_2(\hat{\pi}_{\text{supswr}})$$

$$= E_1 V_2 \left[\sum_{h=1}^{L} Z_h \frac{1}{n_h M_{h0}} \sum_{i=1}^{n_h} \frac{M_{hi} \hat{\pi}_{hi}}{p_{hi}} \right]$$

$$= E_1\left[\sum_{h=1}^{L} Z_h^2 \frac{1}{(n_h M_{h0})^2} \sum_{i=1}^{n_h} \frac{M_{hi}^2}{p_{hi}^2} V_2(\hat{\pi}_{hi})\right]$$

$$= E_1\left[\sum_{h=1}^{L} Z_h^2 \frac{1}{(n_h M_{h0})^2} \times \sum_{i=1}^{n_h} \frac{M_{hi}^2}{p_{hi}^2} \times V_2\left\{\frac{1}{2} + \big((m_{hi11}/m_{hi} - m_{hi00}/m_{hi})B_{hi} + (m_{hi10}/m_{hi} - m_{hi01}/m_{hi})C_{hi}\big)/\big(2\big(B_{hi}^2 + C_{hi}^2\big)\big)\right\}\right]$$

$$= E_1\left[\sum_{h=1}^{L} Z_h^2 \frac{1}{(n_h M_{h0})^2} \times \sum_{i=1}^{n_h} \frac{M_{hi}^2}{p_{hi}^2} \frac{1}{4(B_{hi}^2 + C_{hi}^2)^2} \times \left\{B_{hi}^2 \times V\left(\frac{m_{hi11}}{m_{hi}} - \frac{m_{hi00}}{m_{hi}}\right) + C_{hi}^2 \times V\left(\frac{m_{hi10}}{m_{hi}} - \frac{m_{hi01}}{m_{hi}}\right) + 2B_{hi}C_{hi} \times \mathrm{Cov}\left(\frac{m_{hi11} - m_{hi00}}{m_{hi}}, \frac{m_{hi10} - m_{hi01}}{m_{hi}}\right)\right\}\right].$$

Using the following results from the multinomial distribution, we can arrange the equation $E_1 V_2(\hat{\pi}_{\mathrm{supswr}})$ as follows and prove Theorem 3.2.

$$V(m_{hi11}/m_{hi}) = \theta_{hi11}(1 - \theta_{hi11})/m_{hi};$$
$$\mathrm{Cov}(m_{hi11}/m_{hi}, m_{hi10}/m_{hi}) = -\theta_{hi11}\theta_{hi10}/m_{hi};$$
$$V(m_{hi10}/m_{hi}) = \theta_{hi10}(1 - \theta_{hi10})/m_{hi};$$
$$\mathrm{Cov}(m_{hi10}/m_{hi}, m_{hi01}/m_{hi}) = -\theta_{hi10}\theta_{hi01}/m_{hi};$$
$$V(m_{hi01}/m_{hi}) = \theta_{hi01}(1 - \theta_{hi01})/m_{hi};$$
$$\mathrm{Cov}(m_{hi11}/m_{hi}, m_{hi01}/m_{hi}) = -\theta_{hi11}\theta_{hi01}/m_{hi};$$
$$V(m_{hi00}/m_{hi}) = \theta_{hi00}(1 - \theta_{hi00})/m_{hi};$$
$$\mathrm{Cov}(m_{hi10}/m_{hi}, m_{hi00}/m_{hi}) = -\theta_{hi10}\theta_{hi00}/m_{hi};$$
$$\mathrm{Cov}(m_{hi01}/m_{hi}, m_{hi00}/m_{hi}) = -\theta_{hi01}\theta_{hi00}/m_{hi};$$
$$\mathrm{Cov}(m_{hi11}/m_{hi}, m_{hi00}/m_{hi}) = -\theta_{hi11}\theta_{hi00}/m_{hi}.$$

$$
\begin{aligned}
E_1 V_2(\hat{\pi}_{\text{supswr}}) &= E\left[\sum_{h=1}^{L} Z_h^2 \frac{1}{(n_h M_{h0})^2} \right.\\
&\qquad \times \sum_{i=1}^{n_h} \frac{M_{hi}^2}{p_{hi}^2} \left\{ \frac{B_{hi}^2(E_{hi}+F_{hi}) + C_{hi}^2(G_{hi}+H_{hi})}{4m_{hi}(B_{hi}^2+C_{hi}^2)^2} \right.\\
&\qquad\qquad \left.\left. - \frac{(2\pi_{hi}-1)^2}{4m_{hi}} \right\}\right]\\
&= \sum_{h=1}^{L} Z_h^2 \frac{n_h}{(n_h M_{h0})^2}\\
&\qquad \times \sum_{i=1}^{N_h} p_{hi} \frac{M_{hi}^2}{p_{hi}^2} \left[\frac{B_{hi}^2(E_{hi}+F_{hi}) + C_{hi}^2(G_{hi}+H_{hi})}{4m_{hi}(B_{hi}^2+C_{hi}^2)^2} \right.\\
&\qquad\qquad \left. - \frac{(2\pi_{hi}-1)^2}{4m_{hi}} \right]\\
&= \sum_{h=1}^{L} Z_h^2 \frac{1}{n_h M_{h0}^2}\\
&\qquad \times \sum_{i=1}^{N_h} \frac{M_{hi}^2}{p_{hi}} \frac{1}{4m_{hi}} \left[\frac{B_{hi}^2(E_{hi}+F_{hi}) + C_{hi}^2(G_{hi}+H_{hi})}{(B_{hi}^2+C_{hi}^2)^2} \right.\\
&\qquad\qquad \left. - (2\pi_{hi}-1)^2 \right].
\end{aligned}
$$

□

An unbiased estimator of the variance of $\hat{\pi}_{\text{supswr}}$ is given by

$$
\begin{aligned}
\hat{V}(\hat{\pi}_{\text{supswr}}) = \sum_{h=1}^{L} Z_h^2 &\left[\frac{1}{n_h M_{h0}^2} \sum_{i=1}^{n_h} p_{hi} \left(\frac{M_{hi}\hat{\pi}_{hi}}{p_{hi}} - M_{h0}\hat{\pi}_h \right)^2 \right.\\
&+ \frac{1}{n_h M_{h0}^2} \sum_{i=1}^{n_h} \frac{M_{hi}^2}{p_{hi}} \frac{1}{4(m_{hi}-1)}\\
&\qquad \times \left\{ \frac{B_{hi}^2(E_{hi}+F_{hi}) + C_{hi}^2(G_{hi}+H_{hi})}{(B_{hi}^2+C_{hi}^2)^2} \right.\\
&\qquad\qquad \left.\left. - (2\hat{\pi}_{hi}-1)^2 \right\}\right].
\end{aligned} \tag{3.5}
$$

Meanwhile, if the M_{hi} is known and each unit of n_h PSUs is selected with probability proportional to its size M_{hi}, then the unequal probability $p_{hi} = M_{hi}/M_{h0}$. We call it sampling with probability proportional to size, or *pps* sampling.

In this case, the estimator $\hat{\pi}_h$ of the population proportion of stratum h, π_h, is given by

$$\hat{\pi}_h = \frac{1}{n\,h}\sum_{i=1}^{n_h}\hat{\pi}_{hi}$$
$$= \frac{1}{n_h}\sum_{i=1}^{n_h}\Big[\frac{1}{2} + \big((m_{hi11}/m_{hi} - m_{hi00}/m_{hi})B_{hi} + (m_{hi10}/m_{hi} - m_{hi01}/m_{hi})C_{hi}\big)/\big(2(B_{hi}^2 + C_{hi}^2)\big)\Big], \tag{3.6}$$

and the variance of $\hat{\pi}_{\text{sppswr}}$ and its estimator are respectively

$$V(\hat{\pi}_{\text{sppswr}}) = \sum_{h=1}^{L} Z_h^2\Big[\frac{1}{n_h M_{h0}}\sum_{i=1}^{N_h} M_{hi}(\pi_{hi} - \pi_h)^2 + \frac{1}{n_h M_0}\sum_{i=1}^{N_h} M_{hi}\frac{1}{4m_{hi}} \times \Big(\frac{B_{hi}^2(E_{hi} + F_{hi}) + C_{hi}^2(G_{hi} + H_{hi})}{(B_{hi}^2 + C_{hi}^2)^2} - (2\pi_{hi} - 1)^2\Big)\Big], \tag{3.7}$$

$$\hat{V}(\hat{\pi}_{\text{sppswr}}) = \sum_{h=1}^{L} Z_h^2\Big[\frac{1}{n_h M_{h0}}\sum_{i=1}^{n_h} M_{hi}(\hat{\pi}_{hi} - \hat{\pi}_h)^2 + \frac{1}{n_h M_{h0}}\sum_{i=1}^{n_h} M_{hi}\frac{1}{4(m_{hi} - 1)} \times \Big(\frac{B_{hi}^2(E_{hi} + F_{hi}) + C_{hi}^2(G_{hi} + H_{hi})}{(B_{hi}^2 + C_{hi}^2)^2} - (2\hat{\pi}_{hi} - 1)^2\Big)\Big]. \tag{3.8}$$

3.2 Estimation of a sensitive population proportion by SPPSWOR

Let the population be composed of a number of mutually disjoint L strata of N_h $(h = 1, 2, \ldots, L)$ where each stratum is consisted of N_h clusters of size M_{hi}. The n_h clusters are selected by PPSWOR from the h stratum in which m_{hi} $(i = 1, 2, \ldots, n_h)$ observation units are selected by SRSWR from each cluster.

We estimate the sensitive proportion of population π by applying Abdelfatah et al. (2013) model.

The estimator $\hat{\pi}_h$ of π_h is

$$\hat{\pi}_h = \frac{1}{M_{h0}} \sum_{i=1}^{n_h} \frac{M_{hi}\hat{\pi}_{hi}}{\theta_{hi}}, \tag{3.9}$$

where θ_{hi} is the probability that unit i is included in the sample.

The estimator $\hat{\pi}_{\text{sppswor}}$ of π and its variance are respectively,

$$\hat{\pi}_{\text{sppswor}} = \sum_{h=1}^{L} Z_h \frac{1}{M_{h0}} \sum_{i=1}^{n_h} \frac{M_{hi}\hat{\pi}_{hi}}{\theta_{hi}} \tag{3.10}$$

and

$$\begin{aligned} V(\hat{\pi}_{\text{sppswor}}) = \sum_{h=1}^{L} Z_h^2 \Bigg[& \frac{1}{M_{h0}^2} \sum_{i=1}^{N_h} \sum_{j>i}^{N_h} (\theta_{hi}\theta_{hj} - \theta_{hij}) \left(\frac{M_{hi}\pi_{hi}}{\theta_{hi}} - \frac{M_{hj}\pi_{hj}}{\theta_{hj}} \right)^2 \\ & + \frac{1}{M_{h0}^2} \sum_{i=1}^{N_h} \frac{M_{hi}^2}{\theta_{hi}} \frac{1}{4m_{hi}} \\ & \times \left\{ \frac{B_{hi}^2(E_{hi}+F_{hi}) + C_{hi}^2(G_{hi}+H_{hi})}{(B_{hi}^2 + C_{hi}^2)^2} - (2\pi_{hi} - 1)^2 \right\} \Bigg], \end{aligned} \tag{3.11}$$

where θ_{hij} is the probability that unit i and j are included in the sample.

The variance estimator $\hat{V}(\hat{\pi}_{\text{sppswor}})$ of $V(\hat{\pi}_{\text{sppswor}})$ is given by

$$\begin{aligned} \hat{V}(\hat{\pi}_{\text{sppswor}}) = \sum_{h=1}^{L} Z_h^2 \Bigg[& \frac{1}{M_{h0}^2} \sum_{i=1}^{n_h} \sum_{j>i}^{n_h} \frac{(\theta_{hi}\theta_{hj} - \theta_{hij})}{\theta_{hij}} \left(\frac{M_{hi}\hat{\pi}_{hi}}{\theta_{hi}} - \frac{M_{hj}\hat{\pi}_{hj}}{\theta_{hj}} \right)^2 \\ & + \frac{1}{M_{h0}^2} \sum_{i=1}^{n_h} \frac{M_{hi}^2}{\theta_{hi}} \frac{1}{4(m_{hi}-1)} \\ & \times \left\{ \frac{B_{hi}^2(E_{hi}+F_{hi}) + C_{hi}^2(G_{hi}+H_{hi})}{(B_{hi}^2 + C_{hi}^2)^2} - (2\hat{\pi}_{hi} - 1)^2 \right\} \Bigg], \end{aligned} \tag{3.12}$$

3.3 Estimation of sensitive population proportion by equal stratified two-stage sampling

Let the population be composed of a number of mutually disjoint L strata of N_h $(h = 1, 2, \ldots, L)$ and each stratum is consisted of N_h clusters of size M_{hi}.

The n_h clusters are selected by SRSWR from the hth stratum in which m_{hi} ($i = 1, 2, \ldots, n_h$) observation units are selected by SRSWR from each cluster. The estimator $\hat{\pi}_{\text{swr}}$ of sensitive population proportion π can be obtained by applying Abdelfatah et al. (2013) model.

The estimator $\hat{\pi}_h$ of π_h is

$$\hat{\pi}_h = \frac{N_h}{M_{h0} n_h} \sum_{i=1}^{n_h} M_{hi} \hat{\pi}_{hi}, \tag{3.13}$$

where $\hat{\pi}_{hi}$ is the estimator of sensitive population proportion of ith cluster in stratum h as in (3.1).

Now, the overall estimator $\hat{\pi}_{\text{swr}}$ of π, its variance, and variance estimator are given by

$$\hat{\pi}_{\text{swr}} = \sum_{h=1}^{L} Z_h \frac{N_h}{M_{h0} n_h} \sum_{i=1}^{n_h} M_{hi} \hat{\pi}_{hi}, \tag{3.14}$$

$$\begin{aligned} V(\hat{\pi}_{\text{swr}}) = \sum_{h=1}^{L} Z_h^2 \Bigg[& N_h^2 \frac{1}{n_h M_{h0}^2} \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (M_{hi} \pi_{hi} - \bar{M}_h \pi_h)^2 \\ & + \frac{N_h}{n_h M_{h0}^2} \sum_{i=1}^{N_h} M_{hi}^2 \frac{1}{4 m_{hi}} \\ & \times \left\{ \frac{B_{hi}^2 (E_{hi} + F_{hi}) + C_{hi}^2 (G_{hi} + H_{hi})}{(B_{hi}^2 + C_{hi}^2)^2} \right. \\ & \left. - (2\pi_{hi} - 1)^2 \right\} \Bigg], \end{aligned} \tag{3.15}$$

and

$$\begin{aligned} \hat{V}(\hat{\pi}_{\text{swr}}) = \sum_{h=1}^{L} Z_h^2 \Bigg[& N_h^2 \frac{1}{n_h M_{h0}^2} \frac{1}{n_h - 1} \sum_{i=1}^{n_h} (M_{hi} \hat{\pi}_{hi} - \bar{M}_h \hat{\pi}_h)^2 \\ & + \frac{N_h}{n_h M_{h0}^2} \sum_{i=1}^{n_h} M_{hi}^2 \frac{1}{4(m_{hi} - 1)} \\ & \times \left\{ \frac{B_{hi}^2 (E_{hi} + F_{hi}) + C_{hi}^2 (G_{hi} + H_{hi})}{(B_{hi}^2 + C_{hi}^2)^2} \right. \\ & \left. - (2\hat{\pi}_{hi} - 1)^2 \right\} \Bigg], \end{aligned} \tag{3.16}$$

where $\bar{M}_h = \frac{M_{h0}}{N_h}$.

3.4 A comparison of SPPSWR sampling and equal stratified two-stage sampling

When we set $N-1 \fallingdotseq N$, the difference between the variance (3.7) by SPPSWR and (3.15) by equal two-stage stratified sampling can be approximated as follows.

$$\begin{aligned} V(\hat{\pi}_{\mathrm{swr}}) - V(\hat{\pi}_{\mathrm{sppswr}}) \\ = \sum_{h=1}^{L} Z_h^2 \frac{1}{n_h M_{h0} \bar{M}_h} \Bigg[& \sum_{i=1}^{N} (M_{hi} - \bar{M}_h)^2 \pi_{hi}^2 + \bar{M}_h \sum_{i=1}^{N_h} (M_{hi} - \bar{M}_h) \pi_{hi}^2 \\ & + \sum_{i=1}^{N_h} (M_{hi} - \bar{M}_h)^2 \frac{1}{4 m_{hi}} \\ & \times \left(\frac{B_{hi}^2 (E_{hi} + F_{hi}) + C_{hi}^2 (G_{hi} + H_{hi})}{(B_{hi}^2 + C_{hi}^2)^2} \right. \\ & \left. - (2\pi_{hi} - 1)^2 \right) \\ & + \bar{M}_h \sum_{i=1}^{N_h} (M_{hi} - \bar{M}_h) \frac{1}{4 m_{hi}} \\ & \times \left(\frac{B_{hi}^2 (E_{hi} + F_{hi}) + C_{hi}^2 (G_{hi} + H_{hi})}{(B_{hi}^2 + C_{hi}^2)^2} \right. \\ & \left. - (2\pi_{hi} - 1)^2 \right) \Bigg]. \end{aligned} \tag{3.17}$$

If $M_{hi} = \bar{M}_h = \frac{M_{h0}}{N_h}$ in (3.17), then $V(\hat{\pi}_{\mathrm{swr}}) = V(\hat{\pi}_{\mathrm{sppswr}})$. That is, if the sizes of clusters are equal in stratum h, the selection probabilities are equal to $\frac{1}{N_h}$, and they are equal to the selection probabilities of equal stratified two-stage with replacement sampling. Hence, the efficiency of two methods is equivalent.

If the cluster sizes M_{hi} are different significantly on the right of (3.17), the first term $\sum_{i=1}^{N_h} (M_{hi} - \bar{M}_h)^2 \pi_{hi}^2$ and the third term have value greater than zero and the second $\bar{M}_h \sum_{i=1}^{N_h} (M_{hi} - \bar{M}_h) \pi_{hi}^2$ and fourth term have relatively small value. Therefore, the estimation by ppswr sampling is more profitable than that of equal stratified two-stage with replacement sampling when the cluster sizes M_{hi} of stratum h are unequal.

4 A comparison of *pps* estimation with stratified *pps* estimation in two-stage randomized response model

We compare the efficiency of two estimators of a sensitive population proportion where one is obtained by ppswor sampling and the other by stratified *ppswor* sampling.

The relative efficiency (RE) of two methods is the ratio $V(\hat{\pi}_{\text{ppswor}})/V(\hat{\pi}_{\text{sppswor}})$,

$$\text{RE} = \frac{V(\hat{\pi}_{\text{ppswor}})}{V(\hat{\pi}_{\text{sppswor}})}.$$

Values of RE greater than 1 indicate that the estimator obtained by *sppswor* sampling is more efficient than the estimator obtained by *ppswor* sampling. In order to calculate RE empirically, we assume the population has two strata and

$M_0 = 10{,}000$; $M_1 = 1000$; $M_2 = 2000$; $M_3 = 3000$; $M_4 = 4000$, $m_0 = 1000$; $m_1 = 100$; $m_2 = 200$; $m_3 = 300$; $m_4 = 400$.

Assumption 1. $Z_1 = 0.7$; $Z_2 = 0.3$.

Stratum 1: $M_{10} = 7000$; $M_{11} = 700$; $M_{12} = 1400$; $M_{13} = 2100$; $M_{14} = 2800$, $m_{10} = 700$; $m_{11} = 70$; $m_{12} = 140$; $m_{13} = 210$; $m_{14} = 280$.

Stratum 2: $M_{20} = 3000$; $M_{21} = 300$; $M_{22} = 600$; $M_{23} = 900$; $M_{24} = 1200$, $m_{20} = 300$; $m_{21} = 30$; $m_{22} = 60$; $m_{23} = 90$; $m_{24} = 120$.

Assumption 2. $Z_1 = 0.3$; $Z_2 = 0.7$.

Stratum 1: $M_{10} = 3000$; $M_{11} = 300$; $M_{12} = 600$; $M_{13} = 900$; $M_{14} = 1200$, $m_{10} = 300$; $m_{11} = 30$; $m_{12} = 60$; $m_{13} = 90$; $m_{14} = 120$.

Stratum 2: $M_{20} = 7000$; $M_{21} = 700$; $M_{22} = 1400$; $M_{23} = 2100$; $M_{24} = 2800$, $m_{20} = 700$; $m_{21} = 70$; $m_{22} = 140$; $m_{23} = 210$; $m_{24} = 280$.

Assumption 3. $Z_1 = 0.6$; $Z_2 = 0.4$.

Stratum 1: $M_{10} = 6000$; $M_{11} = 600$; $M_{12} = 1200$; $M_{13} = 1800$; $M_{14} = 2400$, $m_{10} = 600$; $m_{11} = 60$; $m_{12} = 120$; $m_{13} = 180$; $m_{14} = 240$.

Stratum 2: $M_{20} = 4000$; $M_{21} = 400$; $M_{22} = 800$; $M_{23} = 1200$; $M_{24} = 1600$, $m_{20} = 400$; $m_{21} = 40$; $m_{22} = 80$; $m_{23} = 120$; $m_{24} = 160$.

Assumption 4. $Z_1 = 0.4$; $Z_2 = 0.6$.

Stratum 1: $M_{10} = 4000$; $M_{11} = 400$; $M_{12} = 800$; $M_{13} = 1200$; $M_{14} = 1600$, $m_{10} = 400$; $m_{11} = 40$; $m_{12} = 80$; $m_{13} = 120$; $m_{14} = 160$.

Stratum 2: $M_{20} = 6000$; $M_{21} = 600$; $M_{22} = 1200$; $M_{23} = 1800$; $M_{24} = 2400$, $m_{20} = 600$; $m_{21} = 60$; $m_{22} = 120$; $m_{23} = 180$; $m_{24} = 240$.

θ_{hi} and θ_{hij} which are necessary to calculate the variances of $\hat{\pi}_{\text{ppswor}}$ and $\hat{\pi}_{\text{sppswor}}$ can be obtained by

$$\theta_1 = \theta_{11} = \theta_{21} = 0.235; \qquad \theta_2 = \theta_{12} = \theta_{22} = 0.441;$$
$$\theta_3 = \theta_{13} = \theta_{23} = 0.609; \qquad \theta_4 = \theta_{14} = \theta_{24} = 0.715,$$
$$\theta_{12} = \theta_{112} = \theta_{212} = 0.047; \qquad \theta_{13} = \theta_{113} = \theta_{213} = 0.077;$$
$$\theta_{14} = \theta_{114} = \theta_{214} = 0.111,$$
$$\theta_{23} = \theta_{123} = \theta_{223} = 0.161; \qquad \theta_{24} = \theta_{124} = \theta_{224} = 0.233;$$
$$\theta_{34} = \theta_{134} = \theta_{234} = 0.371.$$

We calculate REs by increasing the following values from 0.1 to 0.9 by 0.1.

$$P = P_{11} = P_{12} = P_{13} = P_{14} = P_{21} = P_{22} = P_{23} = P_{24};$$
$$T = T_{11} = T_{12} = T_{13} = T_{14} = T_{21} = T_{22} = T_{23} = T_{24};$$
$$W = W_{11} = W_{12} = W_{13} = W_{14} = W_{21} = W_{22} = W_{23} = W_{24};$$
$$Q = Q_{11} = Q_{12} = Q_{13} = Q_{14} = Q_{21} = Q_{22} = Q_{23} = Q_{24}.$$

The total number of cases are 4,782,969 and among them the number of REs over than 1 are 3,025,887 (63.3%) in case of $Z_1 = 0.7$, $Z_2 = 0.3$ (or $Z_1 = 0.3$, $Z_2 = 0.7$).

Table 4 shows the results of frequencies and percentages according to the values of p among the cases of 3,025,887 which are more efficient than *ppswor* sampling.

The number of REs greater than 1 are 3,045,553 (63.7%) in the cases of $Z_1 = 0.6$, $Z_2 = 0.4$ (or $Z_1 = 0.4$, $Z_2 = 0.6$).

Table 5 shows the results of frequencies and percentage according to the values of p among the 3,045,553 cases that are more efficient than *ppswor* sampling.

Table 4 *The cases of RE according to the values of* p ($Z_1 = 0.7$, $Z_2 = 0.3$ *and* $Z_1 = 0.3$, $Z_2 = 0.7$)

p	Number of cases	RE > 1	%
0.1	531,441	0	0.0
0.2	531,441	1452	0.3
0.3	531,441	146,103	27.5
0.4	531,441	399,024	75.1
0.5	531,441	462,396	87.0
0.6	531,441	489,820	92.2
0.7	531,441	501,220	94.3
0.8	531,441	508,752	95.7
0.9	531,441	517,120	97.3
Total	4,782,969	3,025,887	63.3

Table 5 *The cases of RE according to the values of p* ($Z_1 = 0.6$, $Z_2 = 0.4$ *and* $Z_1 = 0.4$, $Z_2 = 0.6$)

p	Number of cases	RE > 1	%
0.1	531,441	0	0.0
0.2	531,441	1284	0.2
0.3	531,441	156,524	29.5
0.4	531,441	406,019	76.4
0.5	531,441	464,326	87.4
0.6	531,441	490,008	92.2
0.7	531,441	501,360	94.3
0.8	531,441	508,864	95.8
0.9	531,441	517,168	97.3
Total	4,782,969	3,045,553	63.7

5 Conclusions

When the population is composed of the number of different sized clusters, we suggest a two-stage randomized response model with unequal probability sampling by using Abdelfatah et al.'s procedure (2013). We compute the estimate of a sensitive parameter, its variance, and variance estimator for each pps sampling and two-stage equal probability sampling. We extend our model to the case of stratified unequal probability sampling and compute them. Finally, we compare the efficiency of the two estimators, one obtained by unequal probability sampling and the other by stratified unequal probability sampling.

We can see by numerical comparisons that under some conditions, the estimator obtained by sppswor sampling is more efficient than the estimator obtained by *ppswor* sampling about more than 63%.

Acknowledgments

The authors are grateful to the Editor-in-Chief, Professor Francisco Cribari-Neto, Associate Editor and two anonymous learned referees for their valuable comments and suggestions on the original manuscript. This work was supported by Woosuk University.

References

Abdelfatah, S. Mazloum, R. and Singh, S. (2013). Efficient use of a two-stage randomized response procedure. *Brazilian Journal of Probability and Statistics* **27**, 608–617. MR3105047

Mangat, N. S. (1994). An improved randomized response strategy. *Journal of the Royal Statistical Society*, *Ser. B* **56**, 93–95. MR1257798

Mangat, N. S. and Singh, R. (1990). An alternative randomized response procedure. *Biometrika* **77**, 439–442. MR1064823

Odumade, O. and Singh, S. (2009). Efficient use of two decks of cards in randomized response sampling. *Communication in Statistics—Theory and Methods* **38**, 439–446. MR2510796

Warner, S. L. (1965). Randomized response; A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association* **60**, 63–69. Available at http://www.jstor.org/stable/2283137.

G.-S. Lee
Department of Children Welfare
Woosuk University
490 Hujeong-ri, Wanju-gun
Jeonbuk, 565-701
Korea

K.-H. Hong
Department of Computer Science
Dongshin University
252 Daeho-Dong, Naju
Chonnam, 520-714
Korea

J.-M. Kim
Division of Science and Mathematics
University of Minnesota–Morris
Morris, Minnesota 56267
USA
E-mail: jongmink@morris.umn.edu

C.-K. Son
Department of Statistics and Information Science
Dongguk University
Seokjang-dong, Gyeongju
Gyeongbuk, 780-714
Korea