Journal of Statistics Applications & Probability
An International Journal

http://dx.doi.org/10.18576/jsap/110309

Two-stage Adaptive Cluster Sampling to Estimate the Rare Sensitive Parameter under Poisson Distribution

Mahmoud M. Mansour[1,*] *and Enayat M. Abd Elrazik*[2]

[1] Department of Statistics, Mathematics, and Insurance, Benha University, Benha, Egypt
[2] Department of MIS, Yanbu, Taibah University, Medina, Saudi Arabia

Received: 7 Jun. 2021, Revised: 21 Aug. 2021, Accepted: 23 Sep. 2021
Published online: 1 Sep. 2022

Abstract: The estimated mean of the number of individuals holding a rare sensitive attribute (e.g., illegal use of tax evasion, narcotics, domestic abuse, or illicit income) in the population spread over a broad geographical region through traditional sampling structures is difficult to quantify because of the social, political and security circumstances which typically contribute to their concentration in some geographical region. In this paper, The mean of the number of persons possessing a rare sensitive attribute (MNSA) by utilizing the Poisson distribution is estimated using a modified Horvitz-Thompson type of estimator under an adaptive two-stage cluster sampling scheme when an unrelated rare non-sensitive attribute parameter is either known or unknown. The variances of the resultant estimators and their unbiased estimates are expressed. In a small example from the variances and mean squared errors, one sees that the adaptive design with the estimator has the lowest variance compared to the unbiased strategy.

Keywords: Poisson distribution, rare sensitive attribute, unrelated attribute, randomized response, privacy protection, adaptive cluster sampling, Horvitz-Thompson estimator

1 Introduction

Warner [1] presented an excellent method for estimating the proportion of sensitive attributes (such as illegal use of tax evasion, drugs, domestic violence, or illicit income) called a randomized response technique (RRT). Warner's technique was extended by several researchers, including [2,3,4,5,6,7,8,9,10,11] to increase co-operation from the respondents and improve efficiencies of the estimators.
Land et al.[6] introduced a method to estimate the MNSA by utilizing the Poisson distribution in survey sampling. Sing and Tarry [12] suggested an unbiased estimator of the MNSA in the presence of the known proportion of persons possessing a rare unrelated attribute. Lee et al. [9] estimated the MNSA using the Poisson distribution and stratified two-stage sampling and extended the Land et al. [6]. Suman and Singh [13] presented the process for estimating the MNSA when population units are heterogeneous with and without prior information on the supplementary (unrelated rare non-sensitive) characteristic. Sing et al.[14] presented unbiased estimation procedures of the mean total number of persons with a rare sensitive attribute applying a clustered population under two-stage and stratified two-stage sampling schemes.
Adaptive cluster sampling (ACS) is an efficient design for rare and clustered populations (Thompson [15]; Thompson and Seber [16]). Stratified ACS was implemented by [17], and two-stage ACS was developed by [18]. The two-stage ACS is intended to select a fixed number of primary sampling units (PSUs) by simple random sampling without-replacement (SRSWOR) in the first stage and then to select a fixed number of secondary sampling units (SSUs) in the second stage of each chosen PSU, including SRSWOR.
In this paper, Alternative estimation approaches are proposed to estimate the MNSA under the Poisson distribution. We use two-stage ACS sampling schemes. a modified Horvitz-Thompson type of estimator under an adaptive two-stage cluster sampling scheme when the parameter of an unrelated rare non-sensitive attribute is either known or unknown. We express the variances of the estimators and their estimates, when the parameter of the unrelated non-sensitive attribute is either known or unknown.

* Corresponding author e-mail: mahmoud.mansour@fcom.bu.edu.eg

2 Sampling design

For the adaptive cluster sampling designs considered in this paper, the finite population sampling situation, First, let the population be partitioned into N primary sample units PSUs containing $N_1, N_2, \ldots, N_N$ units. Initially, a sample is selected by the traditional two-stage sampling method as follows:

1.A simple random sample (SRS) is selected of size n from the N clusters PUSs.
2.In each selected PSU from the first stage, a large SRS of m_i units is selected from the N_i units.

The remainder of the sample PSUs are selected adaptively by using one condition based on the estimated parameter of the sample units in the PSU. The condition for the additional collection of the neighboring units shall be defined by the interval or C within the range of the interest parameter. The unit i is said to satisfy the condition if $\hat{\theta}_i \in C$. In the examples in this article, a unit satisfies the condition if the estimated value of interest $\theta_i$is greater than or equal to some constant c, that is, $C = \{\hat{\theta}_i : \hat{\theta}_i \geqslant C\}$. When a selected unit satisfies the condition, all units within its neighborhood are added to the sample and observed. Any of these units can fulfill the condition in turn, and some may not. The units in its proximity are also included in the sample for each of these units that satisfy the condition, and so on. Due to the initial selection of unit i consider the list of all the units that are observed under the scheme. When it occurs in a survey, such a set, which could consist of the union of many neighborhoods, would be called a cluster. Within such a cluster is a subcollection of PSUs, termed a network, with the property that selection of any PSU within the network would lead to inclusion in the sample of every other PSU in the network. Any PSU not satisfying the condition but in the neighborhood of one that does is termed an edge unit. Although the selection of any PSU in the network will result in inclusion of all units in the network and all associated edge PSUs, selection of an edge unit will not result in the inclusion of any other units. It is convenient to consider any PSU not satisfying the condition of a network of size one so that, given the y-values, the population may be uniquely partitioned into networks.

3 Estimation procedure of a rare sensitive attribute under adaptive two-stage cluster sampling using a randomized response model

We present the calculation method for the MNSA using a modified Horvitz-Thompson type of estimator under an adaptive two-stage cluster sampling scheme under the adaptive two-stage cluster sampling method. We investigate the situation where an unrelated rare non-sensitive characteristic is known and the situation where it is unknown. In the second stage samples, we gather the elementary units' answers using the randomization device of Sing and Tarray [12].

3.1 When the unrelated rare attribute is known

If the proportion β_u of individuals with the unrelated rare attribute is identified, respondents are invited to use and respond to the randomization device without disclosing the attribute or not. The randomization device consists of a κ_i card deck given to the respondents chosen from the *ith* cluster. The cards have one of the following statements:

1."Do you have the rare sensitive attribute F?," with probability P_{1i};
2."Do you have the rare unrelated attribute U?," with probability P_{2i};
3."Draw one more card," with probability $\mathrm{P}_{3i} = 1 - \mathrm{P}_{1i} - \mathrm{P}_{2i}$.

If the statement (iii) is drawn, the respondent must replicate the above procedure without removing the card. He / she is led to report "No" in the second level, if the statement (iii) is drawn. If κ_i is the total number of cards in the proposed deck, the probability of a "Yes" response is given

$$\varphi_{i0} = \left[\mathrm{P}_{1i}\ \beta_{if} + \mathrm{P}_{2i}\beta_{iu}\right]\left[1 + P_{3i}\kappa_i/(\kappa_i - 1)\right], \tag{1}$$

note that both F and U attributes are very rare in population. As before, assuming that, $m_i \to \infty$, $m_i\varphi_{i0} = \vartheta_{i0} > 0$, as $\varphi_{i0} \to 0$, $m_i\beta_{if} = \vartheta_{if} > 0$, as $\beta_{if} \to 0$, *and* $m_i\beta_{iu} = \vartheta_{iu} > 0$ as $\beta_{iu} \to 0$, where

$$\vartheta_{i0} = \left[\mathrm{P}_{1i}\ \vartheta_{if} + \mathrm{P}_{2i}\vartheta_{iu}\right]\left[1 + P_{3i}\kappa_i/(\kappa_i - 1)\right]. \tag{2}$$

Let $y_{i1}, y_{i2}, \ldots, y_{in}$ be a random sample of m_i observations from the Poisson distribution with parameter ϑ_{i0} from the ith cluster of the population. The likelihood function of the random sample of m_i observations, is given by

$$L = \prod_{j=1}^{m_i} \frac{e^{-\vartheta_{i0}}\ \vartheta_{i0}{}^{y_{ij}}}{y_{ij}!},$$

$$L = e^{-m_i\vartheta_{i0}}\ \vartheta_{i0}{}^{\sum_{j=1}^{m_i} y_{ij}} \prod_{j=1}^{m_i} \frac{1}{y_{ij}!}. \tag{3}$$

From (3), the maximum-likelihood estimator $\widehat{\vartheta}_{if}$ of the MNSA in the ith cluster is

$$\widehat{\vartheta}_{if} = \frac{1}{\mathrm{P}_{1i}} \left[\frac{\sum_{j=1}^{m_i} y_{ij}}{m_i\,(1+P_{3i}\kappa_i/(\kappa_i-1))} - \mathrm{P}_{2i}\vartheta_{iu} \right]. \tag{4}$$

Theorem 3.1. *The estimator $\widehat{\vartheta}_{if}$ is an unbiased estimator of the parameter ϑ_{if}.*

Proof. It is the consequence of the observation that y_{ij} follows a Poisson distribution with parameter ϑ_{i0}. It is the consequence of the observation that $E\,(y_{ij}) = \vartheta_{i0}$. Hence it is easy to conclude that $E\left(\widehat{\vartheta}_{if}\right) = \vartheta_{if}$.

Based on the estimated parameter $\widehat{\vartheta}_{if}$ of the sample units in the cluster i. The condition for the additional collection of the neighboring units shall be defined by the interval or within the range of the interest parameter ϑ_{if}. The unit i is said to satisfy the condition if $\widehat{\vartheta}_{if}\ \varepsilon\ C$. In this article's examples, a unit satisfies the condition if the estimated parameter $\widehat{\vartheta}_{if}$is greater than or equal to some constant c. When a selected unit satisfies the condition, all units within its neighborhood are added to the sample and observed. We estimated the MNSA using a modified Horvitz-Thompson type of estimator under an adaptive two-stage cluster sampling .For sampling designs in which the probability $\alpha_k$that cluster $i$is included in the sample. If the initial cluster is selected by simple random sampling without replacement, the probability that cluster i (inclusion probabilitiesα_k)is included in the sample define as

$$\alpha_k = 1 - \frac{\binom{N-\omega_k}{n_1}}{\binom{N}{n_1}}, \tag{5}$$

where N is the number of units in the population and $\omega_k$number of units on the network that comprises the unit k. If the initial selection is made with replacement, the probability that cluster i (inclusion probabilitiesα_i) is included in the sample define as

$$\alpha_k = 1 - \left(1 - \frac{\omega_k}{N}\right)^{n_1}. \tag{6}$$

For any cluster that does not satisfy the condition, $\omega_k = 1$. Let the indicator variable be 0 if the kth cluster in the sample does not satisfy the condition and was not selected in the initial sample; otherwise, $J_k = 1$. So, a modified Horvitz-Thompson type of estimator of MNSA in the population under adaptive two-stage sampling design is then

$$\widehat{\vartheta}_{fHT} = \frac{1}{N} \sum_{k=1}^{\tau} \frac{\widehat{\vartheta}_{kf}\ J_k}{\alpha_k},$$

where τ is the amount of different PSUs in the survey.

Theorem 3.2. *The estimator $\widehat{\vartheta}_{fHT}$of the MNSA ϑ_f is unbiased.*

Proof. We consider

$$E_1E_2\left(\widehat{\vartheta}_{fHT}\right) = E_1E_2\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\widehat{\vartheta}_{if}\ J_i}{\alpha_i}\right) \tag{7}$$

where E_1 is the expected total number of first-stage selections and E_2 is the expected total number of second-stage selections then

$$E_1E_2\left(\widehat{\vartheta}_{fHT}\right) = E_1\left(\frac{1}{N}\sum_{i=1}^{N}\frac{\vartheta_{if}\ J_i}{\alpha_i}\right) \tag{8}$$

Let $J_i = 1$ if unit i is used in the estimator and $J_i = 0$ otherwise. For any i, J_i is a Bernoulli random variable with expected value α_i *then*

$$E_1E_2\left(\hat{\vartheta}_{fHT}\right) = \left(\frac{1}{N}\sum_{i=1}^{N}\vartheta_{if}\right) = \vartheta_f. \tag{9}$$

Theorem 3.3. *The variance of the estimator* $\widehat{\vartheta}_{fHT}$ *is*

$$Var\left(\widehat{\vartheta}_{fHT}\right) = \frac{1}{N^2}\left[\sum_{k=1}^{\mathrm{K}}\sum_{b=1}^{\mathrm{K}}\vartheta_{kf^*}\ \vartheta_{bf^*}\ \left(\nu_{kb}-\nu_k\nu_b\right)/\left(\nu_k\nu_b\nu_{kb}\right)+\sum_{k=1}^{N}m_k\frac{\Psi_k}{\alpha_k}\right], \tag{10}$$

where

$$\Psi_k = \left(\frac{P_{1k}\ \vartheta_{kf}+P_{2k}\vartheta_{ku}}{P_{1k}^2\left(1+P_{3k}\kappa_k/\left(\kappa_k-1\right)\right)}\right).$$

Proof. Var_1 is the variance of the adaptive first-stage sample and Var_2 the variance over the second-stage sample. The variance of the estimator $\widehat{\vartheta}_{fHT}$ is decomposed as

$$Var\left(\widehat{\vartheta}_{fHT}\right) = Var_1E_2\left(\widehat{\vartheta}_{fHT}\right)+E_1Var_2\left(\widehat{\vartheta}_{fHT}\right). \tag{11}$$

The first term of 11 is

$$Var_1E_2\left(\widehat{\vartheta}_{fHT}\right) = Var_1E_2\left(\frac{1}{N}\sum_{k=1}^{\tau}\frac{\widehat{\vartheta}_{kf}\ J_k}{\alpha_k}\right) = Var_1\left(\frac{1}{N}\sum_{k=1}^{\tau}\frac{\vartheta_{kf}\ J_k}{\alpha_k}\right)$$

Let ζ denote the number of the networks in population, let Φ_i be the set of units comprising the jth network and ω_j number of units on the network that comprises the unit j. The total of estimated parameter in network j will be denoted by $\widehat{\vartheta}_{jf^*} = \sum_{i\in\blacksquare_j}\widehat{\vartheta}_{if}$. The probability α_i that cluster i is included in the estimator is the same for all cluster within a given network j: this probability will be denoted by ν_j. The probability ν_{jh} that the initial sample contains at least one unit in each of the networks j and h is

$$\nu_{jh} = 1-\left\{\binom{N-\omega_j}{n_1}+\binom{N-\omega_h}{n_1}-\binom{N-\omega_j-\omega_h}{n_1}\right\}/\binom{N}{n_1},$$

when the initial sample is selected without replacement and

$$\nu_{jh} = 1-\left\{\left[1-\frac{\omega_j}{N}\right]^{n_1}+\left[1-\frac{\omega_h}{N}\right]^{n_1}-\left[1-\frac{\omega_j+\omega_h}{N}\right]^{n_1}\right\},$$

when the initial sample is selected with replacement. With the convention that $\nu_{jj} = \nu_j$. For any network j, J_j *is* a Bernoulli random variable with expected value $E(J_j) = \nu_j$ and $var(J_j) = \nu_j(1-\nu_j)$ For $j \neq h$, the covariance of the indicator variables is

$$cov\left(\nu_j,\nu_h\right) = E\left(\nu_j\nu_h\right)-E\left(\nu_j\right)E\left(\nu_h\right) = \nu_{jh}-\nu_j\nu_h,$$

thus

$$\begin{aligned} Var_1\left(\frac{1}{N}\sum_{k=1}^{\zeta}\frac{\vartheta_{kf}\ J_k}{\alpha_k}\right) &= \frac{1}{N^2}\sum_{j=1}^{\zeta}\sum_{h=1}^{\zeta}\vartheta_{jf^*}\ \vartheta_{hf^*}\ cov\left(\nu_j,\nu_h\right)/\left(\nu_j\nu_h\right) \\ &= \frac{1}{N^2}\sum_{j=1}^{\zeta}\sum_{h=1}^{\zeta}\vartheta_{jf^*}\ \vartheta_{hf^*}\ \left(\nu_{jh}-\nu_j\nu_h\right)/\left(\nu_j\nu_h\right). \end{aligned}$$

An unbiased estimator of the variance

$$Var_1\left(\frac{1}{N}\sum_{k=1}^{\tau}\frac{\vartheta_{kf}\ J_k}{\alpha_k}\right) == \frac{1}{N^2}\sum_{k=1}^{\mathrm{K}}\sum_{b=1}^{\mathrm{K}}\vartheta_{kf^*}\ \vartheta_{bf^*}\ \left(\nu_{kb}-\nu_k\nu_b\right)/\left(\nu_k\nu_b\nu_{kb}\right). \tag{12}$$

The second term of (11) is

$$
\begin{aligned}
E_1 Var_2\left(\widehat{\vartheta}_{fHT}\right) &= E_1 Var_2\left(\frac{1}{N}\sum_{k=1}^{\tau}\frac{\widehat{\vartheta}_{kf}\,J_k}{\alpha_k}\right)\\
&= E_1\left(\frac{1}{N^2}\sum_{k=1}^{\tau}\frac{J_k Var_2(\widehat{\vartheta}_{kf})}{\alpha_k^2}\right)\\
&= E_1\left[\frac{1}{N^2}\sum_{k=1}^{\tau}\frac{J_k}{\alpha_k^2}\,Var_2\left(\frac{1}{P_{1k}}\left(\frac{\sum_{j=1}^{m_k} y_{kj}}{m_k\left(1+P_{3k}\kappa_k/(\kappa_k-1)\right)}-P_{2k}\vartheta_{ku}\right)\right)\right]\\
&= E_1\left[\frac{1}{N^2}\sum_{k=1}^{\tau}\frac{J_k}{\alpha_k^2}\left(\frac{1}{P_{1k}^2}\left(\frac{\sum_{j=1}^{m_k} Var_2(y_{kj})}{m_k^2\left(1+P_{3k}\kappa_k/(\kappa_k-1)\right)^2}\right)\right)\right]\\
&= E_1\left[\frac{1}{N^2}\sum_{k=1}^{\tau}\frac{J_k}{\alpha_k^2}\left(\frac{1}{P_{1k}^2}\left(\frac{\sum_{j=1}^{m_k}\vartheta_{k0}}{m_k^2\left(1+P_{3k}\kappa_k/(\kappa_k-1)\right)^2}\right)\right)\right]\\
&= E_1\left[\frac{1}{N^2}\sum_{k=1}^{\tau}\frac{J_k}{\alpha_k^2}\left(\frac{1}{P_{1k}^2}\frac{\vartheta_{k0}}{m_k\left(1+P_{3k}\kappa_k/(\kappa_k-1)\right)^2}\right)\right]\\
&= E_1\left[\frac{1}{N^2}\sum_{k=1}^{\tau}\frac{J_k}{\alpha_k^2}\left(\frac{1}{P_{1k}^2}\frac{\left(P_{1k}\,\vartheta_{kf}+P_{2k}\vartheta_{ku}\right)\left(1+P_{3k}\kappa_k/(\kappa_k-1)\right)}{m_k\left(1+P_{3k}\kappa_k/(\kappa_k-1)\right)^2}\right)\right]\\
&= \left[\frac{1}{N^2}\sum_{k=1}^{N} m_k\frac{1}{\alpha_k}\left(\frac{P_{1k}\,\vartheta_{kf}+P_{2k}\vartheta_{ku}}{P_{1k}^2\left(1+P_{3k}\kappa_k/(\kappa_k-1)\right)}\right)\right],
\end{aligned}
\tag{13}
$$

$$
E_1 Var_2\left(\widehat{\vartheta}_{fHT}\right) = \frac{1}{N^2}\sum_{k=1}^{N} m_k\frac{\Psi_k}{\alpha_k}, \tag{14}
$$

where

$$
\Psi_k = \left(\frac{P_{1k}\,\vartheta_{kf}+P_{2k}\vartheta_{ku}}{P_{1k}^2\left(1+P_{3k}\kappa_k/(\kappa_k-1)\right)}\right)
$$

Substituting the values from (12) and (14) into (11), we get the expression of the variance of the estimator $\widehat{\vartheta}_{fHT}$ in (10).

3.2 When the unrelated rare non-sensitive attribute is unknown

Responses are obtained twice from each respondent to estimate the MNSA, while the unrelated uncommon innocuous characteristic is unknown. These randomization devices consist of the decks of κ_i similar cards as described in Section 3.1. Initially, the respondents chosen from the ith cluster are asked to answer "yes" or "no" using the first randomization device based on the

1."Do you have the rare sensitive attribute F?," with probability P_{1i};
2."Do you have the rare unrelated attribute U?," with probability P_{2i};
3."Draw one more card," with probability $P_{3i} = 1 - P_{1i} - P_{2i}$.

As in Section 3.1, the remainder of the process is the same.

Using a second randomization system consisting of the above statements with probabilities T_{1i}, T_{2i}, and T_{3i} instead of P_{1i}, P_{2i}, and P_{3i}, respondents chosen from the ith cluster are again challenged to answer the same queries. As in Section 3.1, the remainder of the process is the same. Based on responses collected using two randomization devices, the probabilities that respondents in the ith cluster answer "yes" are

$$
\varphi_{i1} = \left[P_{1i}\,\beta_{if} + P_{2i}\beta_{iu}\right]\left[1+P_{3i}\kappa_i/(\kappa_i-1)\right], \tag{15}
$$

and

$$\varphi_{i2} = \left[\mathrm{T}_{1i}\ \beta_{if} + \mathrm{T}_{2i}\beta_{iu}\right]\left[1 + T_{3i}\kappa_i/(\kappa_i - 1)\right]. \tag{16}$$

Note that both F and U attributes are very rare in population. As before, assuming that, $m_i \to \infty$, $\varphi_{i1} \to 0$ and $\varphi_{i2} \to 0$, we $\varphi_{i1}\ m_i = \vartheta_{i1} > 0$, *and* $\varphi_{i2} m_i = \vartheta_{i2} > 0$ as $\beta_{iu} \to 0$, where $m_i\beta_{iu} = \vartheta_{iu} > 0$, and $m_i\beta_{if} = \vartheta_{if} > 0$. Subsequently, (15) and (16) we get

$$\vartheta_{i1} = \left[\mathrm{P}_{1i}\ \vartheta_{if} + \mathrm{P}_{2i}\vartheta_{iu}\right]\left[1 + P_{3i}\kappa_i/(\kappa_i - 1)\right], \tag{17}$$

and

$$\vartheta_{i2} = \left[\mathrm{T}_{1i}\ \vartheta_{if} + \mathrm{T}_{2i}\vartheta_{iu}\right]\left[1 + T_{3i}\kappa_i/(\kappa_i - 1)\right]. \tag{18}$$

Likewise, as for (4) and simplifying (16) and (17), we get

$$\frac{1}{m_i}\sum_{j=1}^{m_i} y_{i1j} = \left[\mathrm{P}_{1i}\ \widehat{\vartheta}_{if} + \mathrm{P}_{2i}\widehat{\vartheta}_{iu}\right]\left[1 + P_{3i}\kappa_i/(\kappa_i - 1)\right], \tag{19}$$

$$\frac{1}{m_i}\sum_{j=1}^{m_i} y_{i2j} = \left[\mathrm{T}_{1i}\ \widehat{\vartheta}_{if} + \mathrm{T}_{2i}\widehat{\vartheta}_{iu}\right]\left[1 + T_{3i}\kappa_i/(\kappa_i - 1)\right], \tag{20}$$

where y_{i1j} and y_{i2j} are the first and the second answers of the jth $(j = 1, 2, \ldots, mi)$ respondent in the ith cluster. Solving (19) and (20) , the estimators of ϑ_{if} and ϑ_{iu} are

$$\widehat{\vartheta}_{if2} = \frac{1}{m_i\left(T_{2i}P_{1i} - T_{1i}\ P_{2i}\right)}\left(\frac{T_{2i}\sum_{j=1}^{m_i} y_{i1j}}{1 + P_{3i}\kappa_i/(\kappa_i - 1)} - \frac{P_{2i}\sum_{j=1}^{m_i} y_{i2j}}{1 + T_{3i}\kappa_i/(\kappa_i - 1)}\right), \tag{21}$$

where $T_{2i}P_{1i} \neq T_{1i}\ P_{2i}$

$$\widehat{\vartheta}_{iu2} = \frac{1}{m_i\left(T_{1i}\ P_{2i} - T_{2i}P_{1i}\right)}\left(\frac{T_{1i}\sum_{j=1}^{m_i} y_{i1j}}{1 + P_{3i}\kappa_i/(\kappa_i - 1)} - \frac{P_{1i}\sum_{j=1}^{m_i} y_{i2j}}{1 + T_{3i}\kappa_i/(\kappa_i - 1)}\right), \tag{22}$$

where $T_{2i}P_{1i} \neq T_{1i}\ P_{2i}$

The final estimator of the mean total number of persons having a rare sensitive attribute in the population is then

$$\widehat{\vartheta}_{fHT2} = \frac{1}{N}\sum_{k=1}^{\tau}\frac{\widehat{\vartheta}_{kf2}\ J_k}{\alpha_k}. \tag{23}$$

Theorem 3.4. *The estimator $\widehat{\vartheta}_{fHT2}$of the MNSA attribute ϑ_f is unbiased.*

Proof.

We consider E_1 is the expected total number of first-stage selections and E_2 is the expected total number of second-stage selections then it is easy to conclude that $E\left(\widehat{\vartheta}_{fHT2}\right) = \vartheta_f$.

Theorem 3.5. *The variance of the estimator $\widehat{\vartheta}_{fHT2}$ is*

$$Var\left(\widehat{\vartheta}_{fHT2}\right) = \frac{1}{N^2}\left[\sum_{k=1}^{\mathrm{K}}\sum_{b=1}^{\mathrm{K}}\vartheta_{kf^*}\ \vartheta_{bf^*}\ \left(\nu_{kb} - \nu_k\nu_b\right)/\left(\nu_k\nu_b\nu_{kb}\right) + \sum_{k=1}^{N} m_k\frac{\Phi_k}{\alpha_k}\right], \tag{24}$$

where

$$\Phi_k = \left(P_{1k}H_kT_{2k} + T_{1k}P_{2k}Q_k\right)\vartheta_f\ - \left(P_{2k}H_kT_{2k} + T_{2k}P_{2k}Q_k\right)\vartheta_b - 2T_{2k}P_{2k}(T_{1k}P_{1k}\vartheta_{kf}\ + T_{2k}P_{2k}\vartheta_{ku}\ ),$$

$$H_k = \frac{T_{2k}}{1 + P_{3k}\kappa_i/(\kappa_i - 1)},$$

and

$$Q_k = \frac{P_{2k}}{T_{1k}^2\left\{1 + T_{3k}\kappa_i/(\kappa_i - 1)\right\}}.$$

4 A small example

In this section the sampling techniques are extended in this segment to a tiny community to shed light on adaptive strategies' measurements and features compared to each other and traditional designs. The population consists of just six clusters; a large random sample was taken from each cluster and estimated the MNSA using (4) or (19), the estimated-parameters of which are $\{1,0,1,0,3,5\}$. Both neighboring units are included in each unit (in which there are only one or two). The condition is defined by $C =: \{\widehat{\vartheta}_f : \widehat{\vartheta}_f \geq 2\}$. The initial sample size is $n_1 = 2$.

With the adaptive design in which the initial sample is selected by SRS without replacement, there are $\binom{6}{2} = 15$ possible samples, each having probability 1/15. The resulting observations and the values of each estimator are listed in Table 1.

In this population, the fifth and sixth units, with the $\widehat{\vartheta}_f$ values 3 and 5 , respectively, form a network, and the fourth, fifth, and sixth units, with $\widehat{\vartheta}_f$ values 0, 3, and 5, form a cluster. In the fifth row of the table 1, the first and sixth units, with $\widehat{\vartheta}_f$ values 1 and 5, were selected initially; since $5 > 2$, the double neighbor of the sixth unit, having $\widehat{\vartheta}_f$ values 5, was added to the sample. Since 5 and 3 also exceeds 2, the neighboring units with $\widehat{\vartheta}_f$ values 3,0 is also added to the sample. The computations for the estimators are $\widehat{\vartheta}_{fHT} = (1/0.33 + 5/0.6 + 3/0.6)/6 = 2.72$, in which $\alpha_1 = 1 - \binom{5}{2}/\binom{6}{2} = 0.33$, $\alpha_2 = \alpha_3 = 1 - \binom{4}{2}/\binom{6}{2} = 0.6$. The estimator $\overline{\widehat{\vartheta}_{f1}} = (1+3)/2 = 2$ is obtained by averaging the original units, the classical estimator $\overline{\widehat{\vartheta}_f} = 2.25$ is obtained by averaging all four observations in the sample, and $\overline{\overline{\widehat{\vartheta}_f}} = (1+(5+3+0)/3)/2 = 1.83$.

From the variances and mean squared errors given in the last row of the table, one sees that, for this population, the adaptive design with the estimator $\widehat{\vartheta}_{fHT}$ has the lowest variance compare be the unbiased strategy.

Table 1: The resulting observations and the values of each estimator

networks	$\widehat{\vartheta}_{fHT}$	$\overline{\widehat{\vartheta}_{f1}}$	$\overline{\widehat{\vartheta}_f}$	$\overline{\overline{\widehat{\vartheta}_f}}$
1,0	0.5	0.5	0.5	0.5
1,1	1	1	1	1
1,0	0.5	0.5	0.5	0.5
1,3;0,5	2.75	2	2.25	1.83
1,5;3,0	2.75	3	2.25	1.83
0,1	0.5	0.5	0.5	0.5
0,0	0	0	0	0
0,3;0,5	2.22	1.5	2	1.33
0,5;3	2.22	2.5	2.67	2
1,0	0.5	0.5	0.5	0.5
1,3;0,5	2.75	2	2.25	1.83
1,5;0,3	2.75	3	2.25	1.83
0,3;5	2.22	1.5	2.67	2
0,5;3	2.22	2.5	2.67	2
3,5;0	2.22	4	2.67	2.67
Mean	1.67	1.67	1.65	1.35
Bias	0	0	-0.02	-0.32
MSE	0.97	1.29	0.94	0.58

5 Concluding

It is clear from the estimated average number of persons with a rarely sensitive feature (e.g., illegal usage of tax avoidance, drugs, domestic violence, or illicit income) in the community scattered across a large geographical area through traditional sampling structures is challenging. So we advise, Due to the social, political, and safety factors that usually lead to their concentration in some geographical regions, using two-stage adaptive cluster sampling to estimate the uncommon sensitive parameter under Poisson distribution. Thus, the proposed technique to the estimated average number of persons with a rarely sensitive feature is therefore recommended for its use in practice as an alternative to Land et al. [6], Sing and Tarry [12] , Singh and Suman [14], Suman and Singh [13] and Narjis and Shabbir [**?**] RRT models.

Acknowledgement

The authors are grateful to the anonymous referee for a careful checking of the details and for helpful comments that improved this paper.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] S. L Warner, Randomized response: A survey technique for eliminating evasive answer, Journal of the American Statistical Association,**60(309)** , 63-69 (1965).
[2] B. D Greenberg, A. L Abul-Ela, W. R Simmons, D. G Horvitz, The unrelated question randomized response model: Theoretical framework, Journal of the American Statistical Association, **64(326)**, 520-539 (1969).
[3] L. A Franklin, A comparison of estimators for randomized response sampling with continuous distribution from dichotomous populations, Communications in Statistics, Theory and Methods, **18**, 525-539 (1989).
[4] N. S Mangat, R. Singh, An alternative randomized response procedure, Biometrika,**77(2)** , 439-442 (1990).
[5] J. Kim, M. Elam, A two-stage stratified Warner's Randomized Response Model using optimal allocation, Metrik,**61** , 1-7(2005).
[6] M. Land, S. Singh, S. A Sedory, Estimation of a rare sensitive attribute using Poisson distribution, Statistics, **46(3)** , 351-360 (2012).
[7] G. S Lee, D. Uhm, J. M Kim , Estimation of a rare sensitive attribute in a stratified sample using poisson distribution, Statistics , **47 (3)**, 575-589 (2013).
[8] G. S Lee,K. H Hong, J. M Kim, C. K Son, Estimation of the proportion of a sensitive attribute based on a two-stage randomized response model with stratified unequal probability sampling, Brazilian Journal of Probability and Statistics , **28(3)**, 381-408 (2014).
[9] G.S Lee, K. H Hong, C. K Son, A stratified two-stage unrelated randomized response model for estimating a rare sensitive attribute based on the poisson distribution, Journal of Statistical Theory and Practice , **10(2)**, 239-262 (2016).
[10] G.S Lee,C.K Son, An estimation of a sensitive attribute using adjusted Kuk's randomization device with stratified unequal probability sampling, Communications in Statistics - Theory and Methods, (2020).
[11] M. M Mansour, M. Enayat, Adaptive cluster sampling randomized response model with electronically application, J. Nonlinear Sci. Appl,**14** , 8-14 (2021).
[12] H. P Singh,T. A Tarray, A dexterous randomized response model for estimating a rare sensitive attribute using Poisson distribution, Statistics and Probability Letters, **90**, 42-45 (2014).
[13] S. Suman,G. N Singh, An ameliorated stratified two-stage randomized response model for estimating the rare sensitive parameter under Poisson distribution, STATISTICS,(2019).
[14] G. N Singh,S. Suman, S. Chandraketu, Estimation of a rare sensitive attribute in two-stage sampling using a randomized response model under Poisson distribution, Mathematical Population Studies,(2019).
[15] S. Thompson, Adaptive cluster sampling, Journal of the American Statistical Association, **85**, 1050?1059,(1990).
[16] S. Thompson, Adaptive Sampling, John and Wiley, Sons,(1996).
[17] S. Thompson, Stratified adaptive cluster sampling, Biometrika,**78** , 389-397,(1991).
[18] M. Salehi, G. Seber, Two-stage adaptive cluster sampling, Biometrics, **53**, 959-970,(1997).