Hindawi
Journal of Probability and Statistics
Volume 2022, Article ID 4530120, 14 pages
https://doi.org/10.1155/2022/4530120

Research Article

Mean Estimation of a Sensitive Variable under Nonresponse Using Three-Stage RRT Model in Stratified Two-Phase Sampling

Ronald Onyango, Brian Oduor, and Francis Odundo

Department of Applied Statistics, Financial Mathematics and Actuarial Science, Jaramogi Oginga Odinga University of Science and Technology, Bondo, Kenya

Correspondence should be addressed to Ronald Onyango; assangaronald@gmail.com

Received 20 December 2021; Accepted 2 March 2022; Published 22 April 2022

Academic Editor: Marek T. Malinowski

The present study addresses the problems of mean estimation and nonresponse under the three-stage RRT model. Auxiliary information on an attribute and variable is used to propose a generalized class of exponential ratio-type estimators. Expressions for the bias, mean squared error, and minimum mean squared error for the proposed estimator are derived up to the first degree of approximation. The efficiency of the proposed estimator is studied theoretically and numerically using two real datasets. From the numerical analysis, the proposed generalized class of exponential ratio-type estimators outperforms ordinary mean estimators, usual ratio estimators, and exponential ratio-type estimators. Furthermore, the efficiencies of the mean estimators are observed to decrease with an increase in the sensitivity level of the survey question. As the inverse sampling rate and nonresponse rate go up, so does the efficiency of the mean estimators, which makes them more accurate.

1. Introduction

When conducting a survey, a researcher faces the challenge of estimating the mean in the presence of social desirability and nonresponse. The inability of a survey to collect data from some of the units due to their absence or refusal to participate is referred to as "nonresponse." Nonresponse is a significant issue when the responding and nonresponding units have dissimilar properties. Nonresponse reduces the size of the sample in a survey, increasing the variance of the mean estimate. As a result, the estimator's efficiency suffers, resulting in skewed estimates.

The vast majority of researchers conduct surveys in the hope of gathering reliable data to estimate demographic parameters. However, collecting precise data in a survey about a sensitive subject such as personal income, alcohol consumption, induced abortion, tax evasion, the number of sexual partners, negative website usage, homosexuality, reckless driving, indiscriminate gambling, domestic violence, or illicit drug use, to name a few, is difficult. Correct responses to such sensitive variables are difficult to obtain during personal interviews involving direct questioning of individuals because the respondent's privacy is violated. In reality, most respondents are always hesitant to provide an unvarnished response to a contentious subject for fear of embarrassment or loss of status. As a result, the respondent will either refuse to answer the question or provide an intentionally incorrect response. Warner [1] created the randomized response technique (RRT), which aims to reduce nonresponse rates in surveys with a sensitive variable by keeping respondents anonymous.

The RRT uses a scrambled variable that is independent of the study and auxiliary variables to estimate the mean of a sensitive study variable. The respondent is expected to provide a correct response to the nonsensitive auxiliary variable and a scrambled response to the study variable. In the additive RRT model, the respondent scrambles the genuine answer to a sensitive question (Pollock and Bek [2]). The survey practitioners are unaware of the value-added, but the probability distribution of the scrambled response is assumed to be known. By adding a random number to the correct answer to a sensitive question, a scrambled response is created. The value added is unknown to the survey practitioners, but the probability distribution of the scrambled response is assumed to be known.

Chaudhuri and Mukherjee [3] pioneered the optional RRT model. If a respondent believes that the question is sensitive, the strategy involves giving them the option of responding directly or scrambling. Gupta et al. [4] proposed a one-stage optional RRT model in which the respondent provides a direct response if the question is not sensitive and a scrambled response if it is sensitive. To improve respondent participation and privacy, Gupta et al. [5] proposed a two-stage additive optional RRT model. A predetermined number of respondents t_h are asked to provide a direct response to a sensitive question, while the remaining 1-t_h are asked to provide a scrambled response. However, in order to ensure a high level of privacy and respondent cooperation, the technique requires a high value of t_h.

Mehta [6] proposed a three-stage optional RRT model to encourage respondent cooperation and privacy. In the first stage, a predetermined number of respondents t_h are asked to provide a direct response to a sensitive subject. Another predetermined proportion f_h is asked to scramble their response in the second stage. The remaining proportion$1 - t_h - f_h$ is then given the option of providing a direct or scrambled response. Neeraj and Mehta [7] provided more details on the additive three-stage RRT model.

Several researchers have studied the problem of mean estimation and the RRT model at the same time in the literature. Sousa et al. [8] proposed a ratio estimator of a sensitive variable's population mean in the presence of auxiliary information and a non-optional RRT model. Gupta et al. [9] investigated mean estimation and non-optional RRT in simple random sampling using a generalized mixture estimator. Mushtaq et al. [10] proposed a ratio, regression, and general class of mean estimators of a sensitive variable in stratified two-phase sampling using a non-optional RRT model. Mushtaq et al. [11] proposed a family of estimators in stratified random sampling that use auxiliary information in the presence of a non-optional RRT model. Shahzad et al. [12] proposed a new family of estimators for the mean of a sensitive study variable in simple random sampling using a non-optional RRT model and a single auxiliary variable.

In a survey, nonresponse is accounted for through imputation, weight adjustment processes, and the Hansen and Hurwitz technique [13]. In the presence of missing data, Khalid et al. [14] proposed some estimation procedures for mean estimation using alternative imputation methods under two-occasion successive sampling. In two-occasion successive sampling, Khalid and Singh [15] proposed an alternative imputation method for dealing with the problem of random nonresponse. Khalid and Singh [16] proposed a general class of mean estimators in two-occasion successive sampling under the assumption that the number of nonresponding units follows a discrete probability distribution due to random nonresponse behaviour. The proposed estimators outperformed the existing mean estimators. Shahzad et al. [17] proposed some adapted mean estimators using auxiliary attributes in the presence of nonresponse on a survey variable under stratified random sampling. Zahid et al. [18] addressed the problems of mean estimation, nonresponse, and measurement errors in simple random sampling using a non-optional RRT model. Naeem and Shabbir [19] discussed the issue of mean estimation and nonresponse in the context of two-occasion sampling. Zhang et al. [20] used a one-stage optional RRT model to investigate mean estimation of a sensitive study variable in the presence of nonresponse and measurement errors.

The goal of this study is to address the problem of mean estimation in the presence of nonresponse using a three-stage RRT model in stratified two-phase sampling. Furthermore, the effect of nonresponse and the three-stage RRT model on mean estimation is investigated.

The remaining part of this paper is organised as follows. Section 2 provides a comprehensive overview of the population under consideration. Section 3 examines some of the existing mean estimators in the presence of nonresponse using a three-stage RRT model. Section 4 introduces the proposed generalized class of exponential ratio-type estimators as well as their theoretical bias and MSE properties. Section 5 investigates the proposed estimator's theoretical efficiency. Section 6 examines the numerical performance of the proposed estimators. The study's findings are discussed in Section 7. Section 8 contains a summary of the research.

2. Sampling Strategy and Notations

We consider a finite population $U = \{\Lambda_1, \Lambda_2, \Lambda_3, \ldots, \Lambda_N\}$ of a size N that can be stratified into L homogenous strata with the h^{th} stratum containing N_h, $(h = 1, 2, \ldots, L)$. The sensitive study variable, auxiliary variable, and scrambled response are denoted as Y, X, and Z, respectively. Let Y_{hi}, X_{hi}, and Z_{hi} denote the values of Y, X, and Z, respectively, for the i^{th} value in the h^{th} stratum. Furthermore, let S^2_{Yh}, S^2_{Xh}, and S^2_{Zh} denote the population variances of the survey variable, auxiliary variable, and scrambled response, respectively. Additionally, let S_{XZh} and ρ_{XZh} be the covariance and coefficient of correlation between their subscripts, respectively.

Let τ_{hij} denote the value of an auxiliary attribute τ_j in the h^{th} stratum ($i = 1$, 2, and $j = 1$, 2). If the i^{th} population unit possesses and does not possess an attribute, the auxiliary attribute takes the values of 1 and 0, respectively. Let $A_{hj} = \sum_{i=1}^{N_h} \tau_{hij}$ and $P_{hj} = A_{hj}/N_h$ denote the total number of units with an attribute and the proportion of units with an attribute, respectively, in the h^{th} stratum. Furthermore, let $S^2_{Ph} = 1/N_h - 1 \sum_{i=1}^{N_h} (\tau_{hij} - p_{hj})^2$ be the population variance of an auxiliary attribute in the h^{th} stratum. Let S_{YPh}, S_{XPh} and ρ_{YPh}, ρ_{XPh} denote the population bicovariance and biserial coefficient of correlation between their subscripts, respectively.

In the presence of nonresponse in a survey, the h^{th} stratum population is divided into responding and nonresponding groups of sizes N_{1h} and N_{2h}, respectively. Let $S^2_{Yh(2)}, S^2_{Xh(2)}, S^2_{Xh(2)}$, and $S^2_{Ph(2)}$ denote the population variances of Y, X, Z, and auxiliary attribute for the nonresponding units, respectively. Additionally, let $S_{ZXh(2)}, S_{ZPh(2)}, S_{XPh(2)}$ and $\rho_{ZXh(2)}\, \rho_{ZPh(2)}\, \rho_{XPh(2)}$ denote the population bicovariance and biserial coefficient of correlation between their subscripts for the nonresponding group, respectively.

Recently, several researchers have attempted to improve the efficiency of mean estimators by taking advantage of the availability of known conventional and nonconventional measures of auxiliary variables. Abid et al. [21], Almanjahie et al. [22], and Subhash et al. [23] have used conventional measures of dispersion to propose different mean estimators. Shahzad et al. [24] proposed an exponential-type estimator based on the known median of the study variable. Shahzad et al. [25] used supplemental information on minimum covariance determinant-based quantile to propose a robust regression-type mean estimator under simple random sampling. On the one hand, some of the most common conventional measures of an auxiliary variable are the coefficient of correlation, coefficient of variation, coefficient of skewness, and coefficient of kurtosis.

The coefficient of variation is defined as $C_{Xh} = S_{Xh}/\overline{X}_h$, coefficient of skewness is defined as $\beta_{1h}(x) = N_h \sum_{i=1}^{N} (X_{hi} - \overline{X}_h)^3/(N_h - 1)(N_h - 2)S_{Xh}^3$, and coefficient of kurtosis is defined as

$\beta_{2h}(x) = (N_h(N_h+1)\sum_{i=1}^{N}(X_{hi}-\overline{X}_h)^4/(N_h-1)(N_h-2)(N_h-3)\ S_{Xh}^4)\ -3((N_h-1)/(N_h-2)3(N_h-3)^2)$. On the other hand, nonconventional measures of an auxiliary variable include the midrange value, trimean, quartile deviation, and the Hodges–Lehmann [26] estimator. The midrange is defined as $MR_h(x) = x_{h(1)} + x_{h(Nh)}/2$, where $x_{h(1)}$ is the minimum value and $x_{(Nh)}$ is the maximum value in a dataset. Turkey [27] proposed the trimean, which is defined as $TM_h(x) = Q_{1h}(x) + 2Q_{2h}(x) + Q_{3h}(x)/4$, where $Q_{1h}(x), Q_{2h}(x)$, and $Q_{3h}(x)$ are the first, second, and third quartiles, respectively. The quartile deviation is defined as $QD_h(x) = Q_{3h}(x) - Q_{1h}(x)/2$. The Hodges–Lehmann [26] estimator is defined as $HL_h(x) = \text{Median}(x_{jh} + x_{kh}/2)$, $1 \leq jh \leq kh \leq N$.

Under stratified two-phase sampling, a first phase sample of a certain size n_h' is selected from the population using simple random sampling without replacement (SRSWOR). Thereafter, a second phase sample of size n_h is obtained from the first phase sample using SRSWOR. In the second phase sample n_{1h}, units are observed to respond while the remaining n_{2h} units do not. Let $\overline{x}_h' = (1/n_h')\sum_{i=1}^{n_h'} x_{hi}$ and $p_h' = (a_{hj}/n_h')$ be the sample mean of an auxiliary variable and the proportion of units with an auxiliary attribute, respectively, in the first phase sample. Furthermore, let $\overline{z}_{1h} = (1/n_{1h})\sum_{i=1}^{n_{1h}} z_{hi}$ and $\overline{x}_{1h} = (1/n_{1h})\sum_{i=1}^{n_{1h}} x_{hi}$ be the sample means of Z and X for the responding group in the second phase sample. A subsample of size $r_{2h} = (n_{2h}/k_{2h})$ is drawn from the nonresponding sample, where k_{2h} is the inverse sampling rate. Let $\overline{z}_{2h} = (1/r_{2h})\sum_{i=1}^{r_{2h}} z_{hi}$ and $\overline{x}_{2h} = (1/r_{2h})\sum_{i=1}^{r_{2h}} x_{hi}$ be the subsample means of Z and X, respectively.

Let $p_{1h} = (a_{hj}/n_{1h})$ and $p_{2h} = (a_{hj}/r_{2h})$ denote the proportion of responding units with an auxiliary attribute in the second stage sample and the nonresponding units in the subsample, respectively. The estimates of the population mean for the survey and auxiliary variables in the h^{th} stratum are $\overline{y}_h^* = w_{1h}\overline{y}_{1h} + w_{2h}\overline{y}_{2h}$ and $\overline{x}_h^* = w_{1h}\overline{x}_{1h} + w_{2h}\overline{x}_{2h}$, respectively, where $w_{1h} = (n_{1h}/n_h)$ and $w_{2h} = (n_{2h}/n_h)$. Furthermore, let $p_h^* = w_{1h}p_{1h} + w_{2h}p_{2h}$ represent an estimate of the population proportion possessing an auxiliary attribute in the h^{th} stratum.

3. Some Existing Estimators

The ordinary mean estimator, the usual ratio estimator, and exponential ratio-type estimator are some of the existing estimators in the presence of nonresponse under the three-stage RRT model.

(i) The ordinary mean estimator is defined as

$$t_0 = \sum_{h=1}^{L} w_h \overline{z}_h^*. \quad (1)$$

The variance of the estimator is given as

$$Var(t_0) \cong \sum_{h=1}^{L} W_h^2 B_h^*. \quad (2)$$

(ii) The usual ratio estimator is defined as

$$t_R = \sum_{h=1}^{L} w_h \overline{z}_h^* \frac{\overline{x}_h'}{\overline{x}_h^*}. \quad (3)$$

The bias and mean squared error (MSE) are given as

$$\text{Bias}(t_R) \cong \sum_{h=1}^{L} \frac{W_h}{\overline{X}_h}\left[\frac{9}{8}R_h\left(A_h^* - C_h^*\right) - \left(E_h^* - D_h^*\right)\right],$$

$$MSE(t_R) \cong \sum_{h=1}^{L} W_h^2\left[B_h + R_h^2\left(A_h^* - C_h^*\right) - 2R_h\left(E_h^* - D_h^*\right)\right]. \quad (4)$$

respectively.

(iii) The exponential ratio-type estimator is defined as

$$t_{ER} = \sum_{h=1}^{L} w_h \overline{z}_h^* \left(\frac{\overline{x}_h' - \overline{x}_h^*}{\overline{x}_h' + \overline{x}_h^*}\right). \quad (5)$$

The bias and mean squared error (MSE) are given as

$$Bias(t_{ER}) \cong \sum_{h=1}^{L} \frac{W_h}{2\overline{X}_h}\left[\frac{3}{4}R_h\left(A_h^* - C_h^*\right) - \left(E_h^* - D_h^*\right)\right],$$

$$MSE(t_{ER}) \cong \sum_{h=1}^{L} W_h^2\left[B_h + \frac{1}{4}R_h^2\left(A_h^* - C_h^*\right) - R_h\left(E_h^* - D_h^*\right)\right]. \quad (6)$$

respectively,

where $A_h^* = \theta_h S_{Xh}^2 + \theta_h^* S_{Xh(2)}^2, B_h^* = \theta_h S_{Zh}^2 + \theta_h^* S_{Zh(2)}^2 C_h^* = \theta_h' S_{Xh}^2$, $D_h^* = \theta_h' S_{ZXh}, E_h^* = \theta_h S_{ZXh} + \theta_h^* S_{ZXh(2)}, \theta_h' = (1/n_h' - 1/N_h)$, $\theta_h = (1/n_h - 1/N_h), \theta_h^* = W_h(k_{2h} - 1)/n_h$, and $W_h = N_h/N$

4. Methodology

Various researchers have discussed the problem of mean estimation and nonresponse under non-optional RRT models, one-stage RRT models, and two-stage RRT models in the literature. The problem of mean estimation, however, has been ignored in the three-stage RRT model. Furthermore, there is no literature on the issue of mean estimation in the presence of social desirability bias and nonresponse in stratified two-phase sampling. This study fills a gap in the literature by proposing a generalized class of exponential ratio-type estimators that can be used in the case of nonresponse. It does this by using the three-stage RRT model and auxiliary information.

Neeraj and Mehta [7] assumed that the sensitivity level is known and proposed an additive three-stage RRT model in which a respondent is required to provide a scrambled response defined as

$$Z_{hi} = \begin{cases} Y_{hi}, & \text{with probability } (t_h + 1 - t_h - f_h)(1-\pi_h) \\ Y_{hi} + S_{hi}, & \text{with probability } f_h + (1 - t_h - f_h)\pi_h \end{cases}, \tag{7}$$

where π_h and S_{hi} denote the sensitivity level and scrambling variable, respectively. The scrambling variable has a known mean and variance of 0 and S_{Sh}^2, respectively.

The expectation of the scrambled response under randomization mechanisms is given as

$$\begin{aligned} E_R(Z_{hi}) &= E_R[Y_{hi}(1-\varphi_h) + (Y_{hi} + S_{hi})\varphi_h], \\ E_R(Z_{hi}) &= Y_{hi} + \varphi_h \overline{S}_h, \\ \text{where } \varphi_h &= f_h + \pi_h(1 - t_h - f_h). \end{aligned} \tag{8}$$

The variance of the response variable under randomization mechanisms is given as

$$\begin{aligned} V_R(Z_{hi}) &= V_R(Y_{hi} + \varphi_h S_{hi}), \\ V_R(Z_{hi}) &= \varphi_h^2\left(S_{Sh}^2 + \overline{S}_h^2\right) - \varphi_h^2 \overline{S}_h^2, \\ V_R(Z_{hi}) &= \varphi_h^2 S_{Sh}^2. \end{aligned} \tag{9}$$

The transformed value of the randomized response is given as

$$\hat{y}_{hi} = z_{hi} - \varphi_h \overline{S}_h, \tag{10}$$

with $E_R(\hat{y}_{hi}) = y_{hi}$ and $V_R(\hat{y}_{hi}) = \varphi_h^2 S_{Sh}^2$, where y_{hi} is the true response.

4.1. Modification of HH Technique [13] under the Three-Stage RRT Model. The use of the Hansen and Hurwitz technique [13] in a survey involving a sensitive variable may result in response bias. Also, the respondent may provide an untruthful response to a sensitive question. In this study, the respondent is given the opportunity to provide a scrambled response using the additive three-stage RRT model in the first and second phases of the Hansen and Hurwitz technique [13].

The modified Hansen and Hurwitz [13] technique with an additive three-stage RRT model added is defined as

$$\begin{aligned} \hat{\overline{y}}_h &= w_{1h}\hat{\overline{y}}_{1h} + w_{2h}\hat{\overline{y}}_{2h}, \\ E\left(\hat{\overline{y}}_h\right) &= \overline{Y}_h, \\ \operatorname{var}\left(\hat{\overline{y}}_h\right) &= E_1\left[V_2\left(\hat{\overline{y}}_h\right)\right] + V_1\left[E_2\left(\hat{\overline{y}}_h\right)\right], \\ \operatorname{var}\left(\hat{\overline{y}}_h\right) &= E_1\left[V_2\left(\hat{\overline{y}}_h\right)\right] + V_1\left[E_2\left(\hat{\overline{y}}_h\right)\right], \\ \operatorname{var}\left(\hat{\overline{y}}_h\right) &= \operatorname{var}(\overline{y}_h) + E_1\left[\frac{n_{1h}}{n_h^2}\frac{\varphi_h^2 S_{Sh}^2}{n_{1h}}\right] + E_1\left[\frac{n_{2h}}{n_h^2}k_{2h}\varphi_h^2 S_{Sh}^2\right], \\ \operatorname{var}\left(\hat{\overline{y}}_h\right) &= \operatorname{var}(\overline{y}_h) + \Omega_h, \end{aligned} \tag{11}$$

where $\Omega = (\Phi_h^2 S_{Sh}^2/n_h)(W_{1h} + W_{2h}k_{2h})$ it is the contribution of the three-stage RRT model to the variance of the Hansen and Hurwitz [13] mean estimator.

4.2. Proposed Generalized Class of Exponential Ratio-type Estimators and Their MSE. The proposed mean estimator of a sensitive study variable using the three-stage RRT model and auxiliary information is defined as

$$t_k = \sum_{h=1}^{L} w_h \psi_h \exp\left(\frac{u_h(\overline{x}_h' - \overline{x}_h^*)}{u_h(\overline{x}_h' + \overline{x}_h^*) + 2v_h}\right), \tag{12}$$

where $\psi_h = \overline{z}_h^* + m_{1h}(\overline{x}_h' - \overline{x}_h^*) + m_{2h}(p_h' - p_h^*)$. α_h, and β_h are appropriately chosen constants; u_h and v_h, are either real numbers or the functions of known population parameters of an auxiliary variable.

To obtain the bias and mean squared error (MSE) expressions for the suggested mean estimators, let

$$\begin{aligned} \sigma_{X1h} &= \overline{x}_h' - \overline{X}_h, \\ \sigma_{P1h} &= p_h' - P_h, \\ \sigma_{Ph} &= p_h^* - P_h, \\ \sigma_{Xh} &= \overline{x}_h^* - \overline{X}_h, \\ \sigma_{Xh} &= \overline{x}_h^* - \overline{X}_h, \\ \sigma_{Zh} &= \overline{z}_h^* - \overline{Z}_h. \end{aligned} \tag{13}$$

We take expectations on both sides of equation (13) to obtain

$$E(\sigma_{Zh}) = E(\sigma_{Xh}) = E(\sigma_{X1h}) = E(\sigma_{P1h}) = E(\sigma_{Ph}) = 0. \tag{14}$$

We square both sides of equation (13) and then introduce expectations to obtain

$$E\left(\sigma_{Xh}^2\right) = \theta_h S_{Xh}^2 + \theta_h^* S_{Xh(2)}^2 = A_h^*, \tag{15}$$

$$E\left(\sigma_{Zh}^2\right) = \theta_h S_{Yh}^2 + \theta_h^* S_{Yh(2)}^2 + \Omega_h = B_h^*, \tag{16}$$

$$E\left(\sigma_{X1h}^2\right) = \theta_h' S_{Xh}^2 = C_h^*, \tag{17}$$

$$E\left(\sigma_{X1h}\sigma_{Zh}\right) = \theta_h' S_{ZXh} = D_h^*, \tag{18}$$

$$E\left(\sigma_{Xh}\sigma_{Zh}\right) = \theta_h S_{ZXh} + \theta_h^* S_{ZXh(2)} = E_h^*, \tag{19}$$

$$E\left(\sigma_{Ph}^2\right) = \theta_h S_{ph}^2 + \theta_h^* S_{ph(2)}^2 = F_h^*, \tag{20}$$

$$E\left(\sigma_{P1h}^2\right) = \theta_h' S_{Ph}^2 = G_h^*, \tag{21}$$

$$E\left(\sigma_{Ph}\sigma_{Zh}\right) = \theta_h S_{Zph} + \theta_h^* S_{Zph(2)} = H_h^*, \tag{22}$$

$$E\left(\sigma_{P1h}\sigma_{Zh}\right) = \theta_h' S_{Zph} = I_h^*, \tag{23}$$

$$E\left(\sigma_{Xh}\sigma_{Ph}\right) = \theta_h S_{Xph} + \theta_h^* S_{Xph(2)} = J_h^*, \tag{24}$$

$$E\left(\sigma_{X1h}\sigma_{Xh}\right) = E\left(\sigma_{X1h}\sigma_{Ph}\right) = E\left(\sigma_{X1h}\sigma_{P1h}\right) = \theta_h' S_{XPh} = L_h^*. \tag{25}$$

We substitute equation (13) in (12) and solve using Taylor's approximation while ignoring terms of order greater than two. After that, we subtract the population mean from both sides to get

$$\left(t_k - \overline{Y}\right) = \sum_{h=1}^{L} w_h \left[\begin{array}{c} \frac{1}{2}\gamma_h \overline{Z}_h \sigma_{X1h} - \frac{1}{2}\gamma_h \overline{Z}_h \sigma_{X1h} + \frac{3}{8}\overline{Z}_h \gamma_h^2 \sigma_{Xh}^2 - \frac{3}{8}\overline{Z}_h \gamma_h^2 \sigma_{X1h}^2 \\ +\sigma_{Zh} - \frac{1}{2}\gamma_h \sigma_{Zh}\sigma_{Xh} + \frac{1}{2}\gamma_h \sigma_{Zh}\sigma_{X1h} + m_{1h}\sigma_{X1h} \\ -\frac{1}{2}\gamma_h m_{1h}\sigma_{Xh}\sigma_{X1h} + \frac{1}{2}\gamma_h m_{1h}\sigma_{X1h}^2 - m_{1h}\sigma_{Xh} + \frac{1}{2}\gamma_h m_{1h}\sigma_{Xh}^2 - \frac{1}{2}\gamma_h m_{1h}\sigma_{Xh}\sigma_{X1h} \\ +m_{2h}\sigma_{P1h} - \frac{1}{2}\gamma_h m_{2h}\sigma_{Xh}\sigma_{P1h}\psi + \frac{1}{2}\gamma_h m_{1h}\sigma_{X1h}\sigma_{P1h} - m_{2h}\sigma_{Ph} \\ +\frac{1}{2}\gamma_h m_{2h}\sigma_{Xh}\sigma_{Ph} - \frac{1}{2}\gamma_h m_{2h}\sigma_{X1h}\sigma_{Ph} \end{array}\right], \tag{26}$$

where $\gamma_h = (a_h / a_h \overline{X}_h + b_h)$.

We take expectations on both sides of equation (26) and substitute equations (14)–(25) to obtain an approximation for the bias as

$$\text{Bias}\left(t_k\right) \cong \sum_{h=1}^{L} \frac{W_h \lambda_h^2}{\left[(3/4)\gamma_h \overline{Z}_h \left(A_h^* - C_h^*\right) + m_{1h}\left(A_h^* - C_h^*\right) - \left(E_h^* - D_h^*\right) + m_{2h}\left(J_h^* - L_h^*\right)\right]}. \tag{27}$$

We square both sides of equation (26) and simplify while ignoring terms of order greater than two. Thereafter, we take expectations on both sides and substitute equations (15)–(25) to obtain an approximation for the MSE as

$$MSE\left(t_k\right) \cong \sum_{h=1}^{L} W_h^2 \left[B_h^* + \Delta_{1h} + m_{1h}^2 \Delta_{2h} + m_{2h}^2 \Delta_{3h} + m_{2h}\Delta_{4h} + m_{1h}\Delta_{5h} + 2m_{1h}m_{2h}\Delta_{6h}\right],$$

$$\text{where } \Delta_{1h} = \frac{1}{4}\lambda_h^2 \overline{Y}_h^2 \left(A_h^* - C_h^*\right) - \lambda_h \overline{Y}_h \left(E_h^* - D_h^*\right),$$

$$\Delta_{2h} = \left(A_h^* - C_h^*\right),$$

$$\Delta_{3h} = \left(F_h^* - G_h^*\right),$$

$$\Delta_{4h} = \overline{Z}_h \lambda_h \left(J_h^* - L_h^*\right) - 2\left(H_h^* - I_h^*\right),$$

$$\Delta_{2h} = (A_h^* - C_h^*),$$
$$\Delta_{3h} = (F_h^* - G_h^*),$$
$$\Delta_{4h} = \overline{Z}_h\lambda_h(J_h^* - L_h^*) - 2(H_h^* - I_h^*),$$
$$\Delta_{5h} = \overline{Z}_h\lambda_h(A_h^* - C_h^*) - 2(E_h^* - D_h^*),$$
$$\Delta_{6h} = (J_h^* - L_h^*). \tag{28}$$

We differentiate equation (28) partially with respect to m_{1h} and m_{2h} and then equate to zero to obtain

$$m_{1h}^{(opt)} = \frac{\Delta_{4h}\Delta_{6h} - \Delta_{5h}\Delta_{3h}}{2\left(\Delta_{2h}\Delta_{3h} - \Delta_{6h}^2\right)},$$
$$m_{2h}^{(opt)} = \frac{\Delta_{5h}\Delta_{6h} - \Delta_{4h}\Delta_{2h}}{2\left(\Delta_{2h}\Delta_{3h} - \Delta_{6h}^2\right)}. \tag{29}$$

We substitute equation (29) in (28) to obtain the minimum MSE as

$$MSE(t_k)_{\min} \cong \sum_{h=1}^{L} W_h^2\left[B_h^* + \Delta_{1h} - \frac{\Delta_{4h}^2}{4\Delta_{3h}} - \frac{\left(\Delta_{5h}\Delta_{3h} - \Delta_{4h}\Delta_{6h}\right)^2}{4\Delta_{3h}\left(\Delta_{2h}\Delta_{3h} - \Delta_{6h}^2\right)}\right]. \tag{30}$$

4.3. Members of the Family of Proposed Generalized Class of Exponential Ratio-Type Estimators. Members of the proposed generalized class of exponential ratio-type estimators can be obtained by making appropriate choices of u_h andv_h.

(i) Putting $u_h = 1$ and $v_h = 0$, we get

$$t_1 = \sum_{h=1}^{L} w_h\psi_h \exp\left(\frac{(\overline{x}_h' - \overline{x}_h)}{(\overline{x}_h' + \overline{x}_h)}\right). \tag{31}$$

(ii) Putting $u_h = 1$ and $v_h = C_{Xh}$, we get

$$t_2 = \sum_{h=1}^{L} w_h\psi_h exp\left(\frac{C_{Xh}(\overline{x}_h' - \overline{x}_h^*)}{C_{Xh}(\overline{x}_h' + \overline{x}_h^*) + 2\rho_{XYh}}\right). \tag{32}$$

(iii) Putting $u_h = C_{Xh}$ and $v_h = \rho_{XYh}$, we get

$$t_3 = \sum_{h=1}^{L} w_h\psi_h \exp\left(\frac{C_{Xh}(\overline{x}_h' - \overline{x}_h^*)}{C_{Xh}(\overline{x}_h' + \overline{x}_h^*) + 2\rho_{XYh}}\right). \tag{33}$$

(iv) Putting $u_h = \beta_{1h}(x)$ and $v_h = \rho_{XYh}$, we get

$$t_4 = \sum_{h=1}^{L} w_h\psi_h \exp\left(\frac{\beta_{1h}(x)(\overline{x}_h' - \overline{x}_h^*)}{\beta_{1h}(x)(\overline{x}_h' + \overline{x}_h^*) + 2\rho_{XYh}}\right). \tag{34}$$

(v) Putting $u_h = \beta_{2h}(x)$ and $v_h = \beta_{1h}(x)$, we get

$$t_5 = \sum_{h=1}^{L} w_h\psi_h \exp\left(\frac{\beta_{2h}(x)(\overline{x}_h' - \overline{x}_h^*)}{\beta_{2h}(x)(\overline{x}_h' + \overline{x}_h^*) + 2\beta_{1h}(x)}\right). \tag{35}$$

(vi) Putting $u_h = QD_h(x)$ and $v_h = TM_h(x)$, we get

$$t_6 = \sum_{h=1}^{L} w_h\psi_h \exp\left(\frac{QD_h(x)(\overline{x}_h' - \overline{x}_h^*)}{QD_h(x)(\overline{x}_h' + \overline{x}_h^*) + 2TM_h(x)}\right). \tag{36}$$

(vii) Putting $u_h = QD_h(x)$ and $v_h = MR_h(x)$, we get

$$t_7 = \sum_{h=1}^{L} w_h\psi_h \exp\left(\frac{QD_h(x)(\overline{x}_h' - \overline{x}_h^*)}{QD_h(x)(\overline{x}_h' + \overline{x}_h^*) + 2MR_h(x)}\right). \tag{37}$$

(viii) Putting $u_h = HL_h(x)$ and $v_h = TM_h(x)$, we get

$$t_8 = \sum_{h=1}^{L} w_h\psi_h \exp\left(\frac{HL_h(x)(\overline{x}_h' - \overline{x}_h^*)}{HL_h(x)(\overline{x}_h' + \overline{x}_h^*) + 2TM_h(x)}\right). \tag{38}$$

(ix) Putting $u_h = \rho_{XYh}$ and $v_h = QD_h(x)$, we get

$$t_9 = \sum_{h=1}^{L} w_h\psi_h \exp\left(\frac{\rho_{XYh}(\overline{x}_h' - \overline{x}_h^*)}{\rho_{XYh}(\overline{x}_h' + \overline{x}_h^*) + 2QD_h(x)}\right). \tag{39}$$

(x) Putting $u_h = 1$ and $v_h = \rho_{XYh}$, we get

$$t_{10} = \sum_{h=1}^{L} w_h\psi_h \exp\left(\frac{(\overline{x}_h' - \overline{x}_h^*)}{(\overline{x}_h' + \overline{x}_h^*) + 2\rho_{XYh}}\right). \tag{40}$$

(xi) Putting $u_h = 1$ and $v_h = QD_h(x)$, we get

$$t_{11} = \sum_{h=1}^{L} w_h\psi_h \exp\left(\frac{(\overline{x}_h' - \overline{x}_h^*)}{(\overline{x}_h' + \overline{x}_h^*) + 2QD_h(x)}\right). \tag{41}$$

The bias and mean squared error (MSE) expressions for the special cases of the proposed estimators are obtained by substituting appropriate values of m_{1h} and m_{2h} in (27) and (30) respectively.

5. Theoretical Comparison

In this section, the performance of the proposed estimator is compared theoretically to other existing mean estimators.

Condition 1. From equations (2) and (43), $\mathrm{MSE}(t_k)_{\min} < \mathrm{Var}(t_0)$ when

$$\left[\Delta_{1h} - \frac{\Delta_{4h}^2}{4\Delta_{3h}} - \frac{\left(\Delta_{5h}\Delta_{3h} - \Delta_{4h}\Delta_{6h}\right)^2}{4\Delta_{3h}\left(\Delta_{2h}\Delta_{3h} - \Delta_{6h}^2\right)}\right] < 0. \quad (42)$$

Condition 2. From equations (5) and (43), $\mathrm{MSE}(t_k)_{\min} < \mathrm{MSE}(t_R)$ when

$$\left[\Delta_{1h} - \frac{\Delta_{4h}^2}{4\Delta_{3h}} - \frac{\left(\Delta_{5h}\Delta_{3h} - \Delta_{4h}\Delta_{6h}\right)^2}{4\Delta_{3h}\left(\Delta_{2h}\Delta_{3h} - \Delta_{6h}^2\right)} - R_h^2\left(A_h^* - C_h^*\right) + 2R_h\left(E_h^* - D_h^*\right)\right] < 0. \quad (43)$$

Condition 3. From equations (8) and (43), $\mathrm{MSE}(t_k)_{\min} < \mathrm{MSE}(t_{ER})$ when

$$\left[\Delta_{1h} - \frac{\Delta_{4h}^2}{4\Delta_{3h}} - \frac{\left(\Delta_{5h}\Delta_{3h} - \Delta_{4h}\Delta_{6h}\right)^2}{4\Delta_{3h}\left(\Delta_{2h}\Delta_{3h} - \Delta_{6h}^2\right)} - \frac{1}{4}R_h^2\left(A_h^* - C_h^*\right) - R_h\left(E_h^* - D_h^*\right)\right] < 0. \quad (44)$$

These three conditions are always true. Therefore, the proposed generalized class of exponential ratio-type estimators performs better than other existing mean estimators.

6. Application

A numerical study is conducted to compare the performance of the proposed generalized class of exponential ratio-type estimators to the performance of existing mean estimators. Nonresponse and the three-stage RRT model's effects on mean estimation are also investigated. The COVID-19 global pandemic (http//www.worldometer.info) and Rosner [28] datasets are used. The R programming language is used for data simulation and coding. The proposed estimators' efficiency is compared to that of other estimators using the percent relative efficiency (PRE) approach. To calculate the PREs of the mean estimators,

$$\mathrm{PRE} = \frac{Var(t_0)}{\mathrm{MSE}(t_k)}, \quad (45)$$

where $k = R, ER, 1, 2, \ldots, 11$. The estimator with the highest PRE when compared to the ordinary mean estimator is thought to be more efficient than the others. The PREs are also calculated when the sensitivity level is set to 20% and 80%, respectively.

The following is a description of the data used:

6.1. Population I: COVID-19 Global Pandemic Data. The dataset covers the COVID-19 global pandemic (http//www.worldometer.info) from January 3rd, 2020 to September 17th, 2021. The data are divided into six categories based on World Health Organisation (WHO) regions; African region ($N_1 = 31200$), the American region ($N_2 = 34944$), the Eastern Mediterranean region ($N_3 = 13728$), the European region ($N_4 = 38688$), the South-East Asia region ($N_5 = 6864$), and the Western Pacific region ($N_6 = 21840$).

The auxiliary and study variables are the number of new cases and deaths on a given day, respectively. The auxiliary attribute is the number of new deaths with a value of less than one. A scrambled variable with a mean of 0 and a variance of 2 is generated for each unit in the dataset and used to calculate the scrambled response. Tables 1 and 2 show the population parameter for the responding and nonresponding units, respectively.

6.2. Population II: Rosner [28]. The population is divided into two strata; $N_1 = 480$ and $N_2 = 174$, with forced expiratory volume as the study variable, age (in years) as an auxiliary variable, and gender (Male = 1, Female = 0) as an auxiliary attribute. Furthermore, smoking (Yes = 1, No = 0) is chosen as the scrambling variable and used in the generation of the scrambled response. Tables 3 and 4 show the population parameter for the responding and nonresponding units, respectively.

7. Results and Discussion

Table 5 summarizes the results for the PREs of various mean estimators for population I. The PRE values decrease as the sensitivity level of the survey question increases. For t_5, for example, the value of PRE at 20% nonresponse and $k_{2h} = 2$ is 181.0869 and 181.0735 at 0.2 and 0.8 sensitivity levels, respectively. Furthermore, the values for PREs are found to increase as the inverse sampling rate and nonresponse rate increase. The proposed estimator t_7 has the best PRE of all the estimators that were looked at in this study.

Figures 1–4 show PRE plots for various mean estimators. As the inverse sampling rates rise, the values of PREs for the mean estimators get larger. Generally, the proposed estimators perform better than other existing mean estimators.

Table 6 summarizes the PREs of various mean estimators for population II. From the table, PREs are found to increase in value as the inverse sampling rate and nonresponse rate increase. Furthermore, the values of PREs are found to increase as the sensitivity level of the survey question increases. For example, at 30% nonresponse rate and $k_{2h} = 8$, PRE for t_3 is 155.2214 and 151.2631 at 0.2 and 0.8 sensitivity levels, respectively.

Table 1: Parameters for population I.

Parameter	Stratum 1	Stratum 2	Stratum 3	Stratum 4	Stratum 5	Stratum 6
$\overline{X}_h$	188.9035	2502.012	1120.151	1757.061	6175.008	356.2095
$\overline{Y}_h$	4.543181	61.90972	20.51225	33.79095	97.12205	4.833472
S^2_{Xh}	1094471	187408859	8526375	24712119	817189958	318940
S^2_{Zh}	926.4621	76639.99	2937.237	11588.58	145353	849.8079
S^2_{Ph}	0.2017896	0.2328431	0.2253055	0.2467874	0.247146	0.1323922
ρ_{XZh}	0.8171398	0.7944946	0.834325	0.6559524	0.8679977	0.7237861
ρ_{XPh}	−0.2608673	−0.2379639	−0.265470	−0.2982271	−0.239344	−0.4403104
ρ_{ZPh}	−0.2386865	−0.2924192	−0.272994	−0.2802833	−0.2839612	−0.3832064

Table 2: Parameters for nonresponding units for population I.

Nonresponse rate (%)	Stratum	$S^2_{Xh(2)}$	$S^2_{Ph(2)}$	$S^2_{Zh(2)}$	$S^2_{XZh(2)}$	$S^2_{XPh(2)}$	$S^2_{ZPh(2)}$
20	1	989315.2	0.2050127	763.9706	0.7535727	−0.2691771	−0.2557058
	2	199477087	0.2343414	80842.35	0.8166929	−0.229914	−0.2876454
	3	8176575	0.2249875	2718.171	0.8154184	−0.2564837	−0.2718289
	4	25298233	0.2462478	11741.33	0.6713044	−0.2984032	−0.2839888
	5	708141536	0.2460244	96127.52	0.8765226	−0.2487161	−0.3082972
	6	708141536	0.1284571	857.4592	0.7296066	−0.4390315	−0.402489
30	1	1071816	0.2047748	781.3434	0.7820729	−0.2610072	−0.2520072
	2	206269098	0.2334903	81677.25	0.7969911	−0.2297942	−0.2881574
	3	8525833	0.2250371	2856.095	0.8312795	−0.2655035	−0.2720948
	4	24546992	0.2462213	11427.83	0.6609692	−0.3004036	−0.2837976
	5	681811147	0.2462587	128890.3	0.8571617	−0.2454178	−0.28936
	6	681811147	0.1297743	849.4515	0.7375239	−0.4392996	−0.4083239

Table 3: Parameters for population II.

Parameter	Stratum 1	Stratum 2
$\overline{X}_h$	8.558333	13.71839
$\overline{Y}_h$	2.363715	3.763615
S^2_{Xh}	3.604106	3.301741
S^2_{Zh}	0.5254207	0.7556429
S^2_{Ph}	0.2503653	0.2511461
ρ_{XZh}	0.7239923	0.3619965
ρ_{XPh}	0.02999931	0.07201403
ρ_{ZPh}	0.08365375	0.4809902

Table 4: Parameters for nonresponding units for population II.

Nonresponse rate (%)	Stratum	$S^2_{Xh(2)}$	$S^2_{Ph(2)}$	$S^2_{Zh(2)}$	$S^2_{XZh(2)}$	$S^2_{XPh(2)}$	$S^2_{ZPh(2)}$
20	1	3.803509	0.2486842	0.5833481	0.7305239	0.02164662	0.1079512
	2	4.198319	0.2319328	0.4723521	0.2932356	−0.2094929	0.2899405
30	1	3.746503	0.2513112	0.5804701	0.7363216	0.07657269	0.1124887
	2	4.015239	0.2394775	0.6125016	0.3607974	−0.0199814	0.3836556

PRE plots for different mean estimators for population II are shown in Figures 5–8. PREs are found to increase in value as nonresponse rates and inverse sampling rates increase. The proposed generalized class of exponential ratio-type estimators has higher PREs than existing mean estimators.

Table 5: PREs of different mean estimators for population I.

Estimator	π_h	20% nonresponse k_{2h} 2	4	8	30% nonresponse k_{2h} 2	4	8
t_0	0.2	100	100	100	100	100	100
	0.8	100	100	100	100	100	100
t_R	0.2	115.3888	102.5616	95.65319	112.2147	101.5183	96.6019
	0.8	115.3871	102.5614	95.65344	112.2136	101.5182	96.6022
t_{ER}	0.2	161.3319	172.3874	180.4193	164.6804	175.0072	181.0740
	0.8	161.3229	172.3785	180.4107	164.6716	174.9989	181.0664
t_1	0.2	181.0869	208.5838	230.8077	188.3997	214.4384	231.1358
	0.8	181.0735	208.5675	230.7898	188.3860	214.4230	231.120
t_2	0.2	181.1115	208.712	231.0598	188.4469	214.5998	231.3967
	0.8	181.0981	208.6957	231.0418	188.4332	214.5843	231.3808
t_3	0.2	181.0876	208.5874	230.8149	188.4010	214.4431	231.1434
	0.8	181.0741	208.5711	230.797	188.3873	214.4277	231.1275
t_4	0.2	168.5379	191.8925	210.9381	174.6901	196.8989	211.2498
	0.8	168.5273	191.8798	210.9242	174.6794	196.8869	211.2375
t_5	0.2	181.0869	208.5838	229.9411	188.3997	214.4384	229.4915
	0.8	181.0735	208.5675	229.9234	188.386	214.4230	229.4760
t_6	0.2	181.0893	208.5968	230.8334	188.4044	214.4549	231.1625
	0.8	181.0759	208.5805	230.8155	188.3907	214.4395	231.1466
t_7	0.2	180.1932	210.048	235.4472	187.9831	216.7631	236.0666
	0.8	180.1800	210.0314	235.4283	187.9696	216.7473	236.0498
t_8	0.2	181.0862	208.583	230.8069	188.3990	214.4377	231.1351
	0.8	181.0728	208.5667	230.7890	188.3853	214.4223	231.1192
t_9	0.2	179.9839	207.0395	228.8866	187.1721	212.7807	229.1884
	0.8	179.9707	207.0236	228.8691	187.1587	212.7656	229.1729
t_{10}	0.2	181.0903	208.6023	230.8443	188.4064	214.4618	231.1738
	0.8	181.0769	208.586	230.8264	188.3927	214.4464	231.1579
t_{11}	0.2	180.2022	207.3365	229.2469	187.4102	213.0907	229.5443
	0.8	180.189	207.3205	229.2293	187.3967	213.0756	229.5287

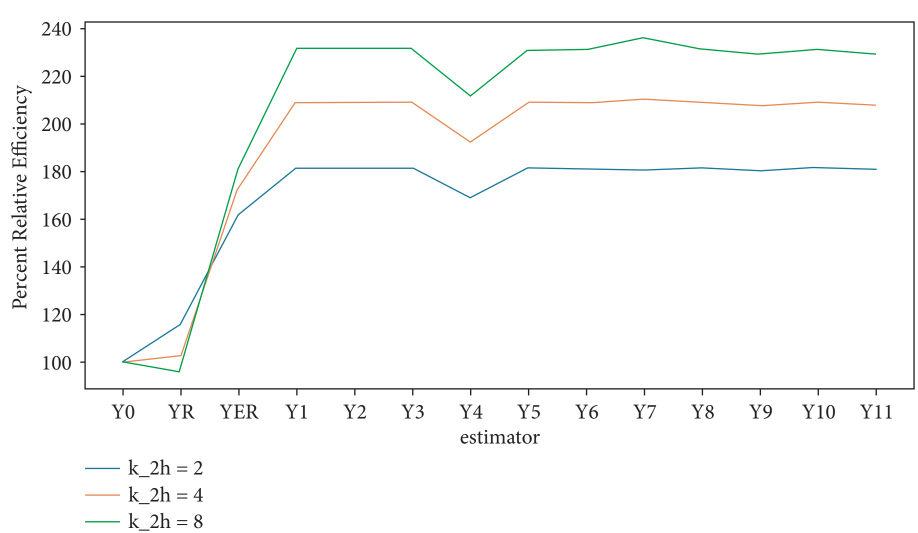

Figure 1: Plot of PREs at 20% nonresponse and 0.2 sensitivity level for population I.

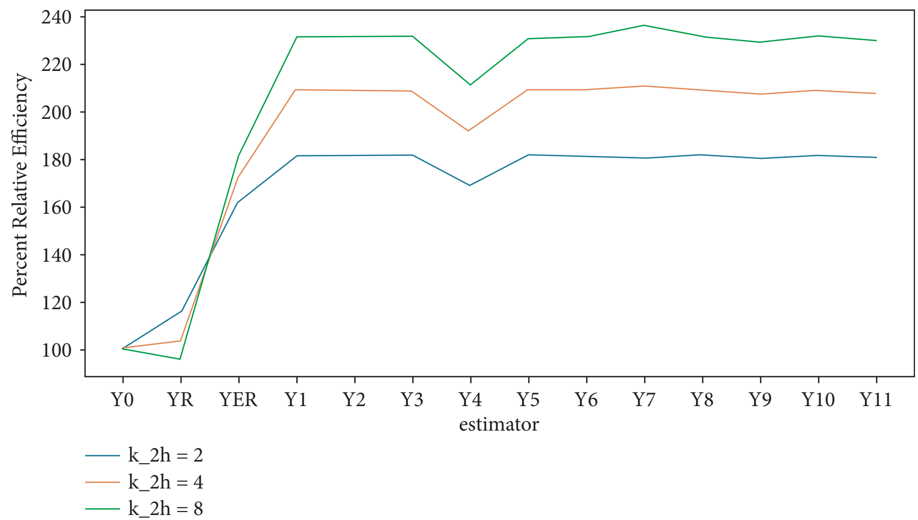

Figure 2: Plot of PREs at 20% nonresponse and 0.8 sensitivity level for population I.

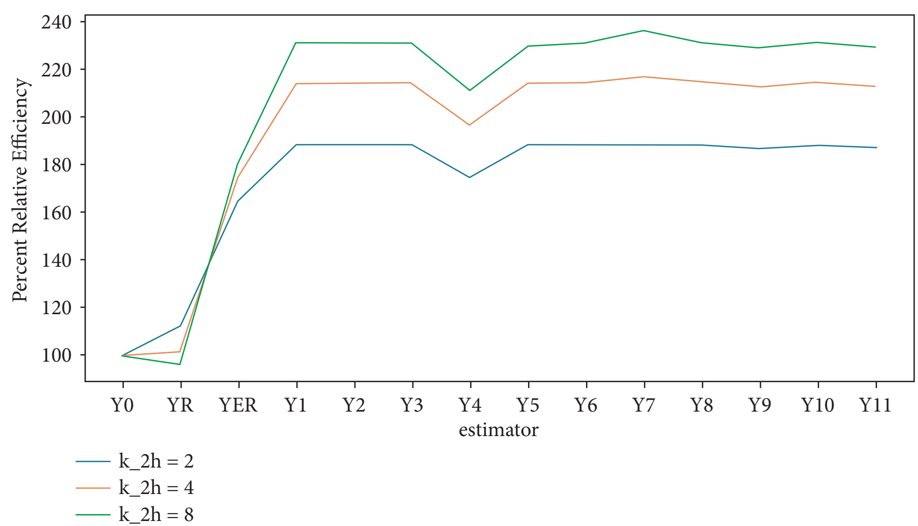

Figure 3: Plot of PREs at 30% nonresponse and 0.2 sensitivity level for population I.

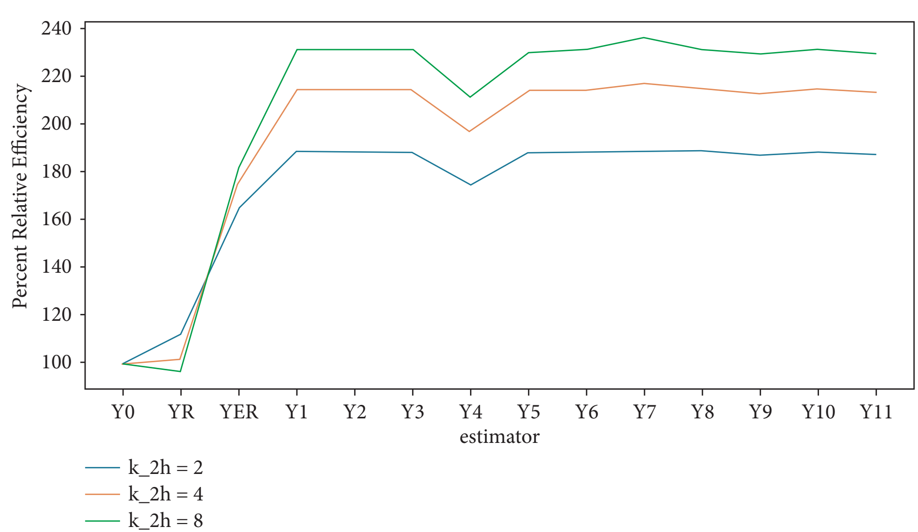

Figure 4: Plot of PREs at 30% nonresponse and 0.8 sensitivity level for population I.

Table 6: PREs of different mean estimators for population II.

	π_h	20% nonresponse k_{2h} 2	4	8	30% nonresponse k_{2h} 2	4	8
t_0	0.2	100	100	100	100	100	100
	0.8	100	100	100	100	100	100
t_R	0.2	127.2439	134.9012	140.8874	132.3932	142.3453	148.7683
	0.8	124.743	132.0578	137.8932	129.5512	139.1108	145.4079
t_{ER}	0.2	122.2609	129.041	134.2996	125.1812	132.5172	137.133
	0.8	120.2903	126.7698	131.8971	123.0822	130.1927	134.7617
t_1	0.2	131.2392	142.4024	151.5345	135.5274	147.4847	155.362
	0.8	128.2900	138.7723	147.4977	132.3433	143.7370	151.3903
t_2	0.2	131.2432	142.4159	151.5557	135.5190	147.4734	155.3484
	0.8	128.2935	138.7843	147.5168	132.3358	143.7269	151.378
t_3	0.2	131.3257	142.6600	151.9334	135.4225	147.3606	155.2214
	0.8	128.3666	139.0018	147.8555	132.2500	143.6256	151.2631
t_4	0.2	131.8164	143.9719	153.8973	135.0548	147.0131	154.8764
	0.8	128.8007	140.1695	149.6149	131.9230	143.3136	150.9510
t_5	0.2	131.2392	142.4024	151.5345	135.5274	147.4847	155.3620
	0.8	128.2900	138.7723	147.4977	132.3433	143.7370	151.3903
t_6	0.2	131.3706	142.7725	152.1017	135.4173	147.3736	155.2473
	0.8	128.4064	139.102	148.0064	132.2454	143.6373	151.2865
t_7	0.2	131.3901	142.8146	152.1626	135.4314	147.4051	155.2916
	0.8	128.4236	139.1394	148.0610	132.2580	143.6656	151.3266
t_8	0.2	131.2759	142.5122	151.7049	135.4828	147.432	155.3024
	0.8	128.3225	138.8701	147.6506	132.3036	143.6897	151.3364
t_9	0.2	131.4914	143.0900	152.5798	135.3591	147.3393	155.2284
	0.8	128.5132	139.3848	148.4349	132.1936	143.6065	151.2694
t_{10}	0.2	131.2487	142.4368	151.5894	135.5019	147.4493	155.3188
	0.8	128.2984	138.8029	147.5469	132.3206	143.7052	151.3512
t_{11}	0.2	131.3516	142.7266	152.0335	135.4151	147.3612	155.2275
	0.8	128.3895	139.0611	147.9453	132.2434	143.6262	151.2686

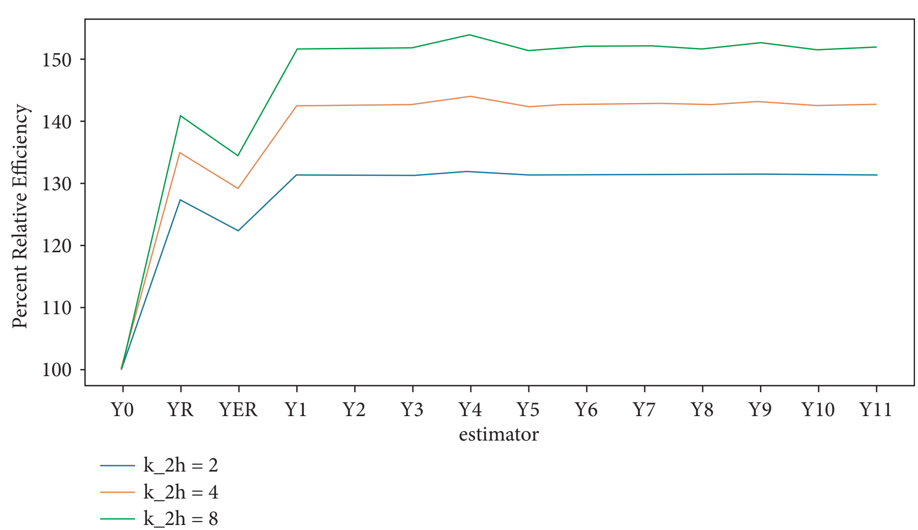

Figure 5: Plot of PREs at 20% nonresponse and 0.2 sensitivity level for population II.

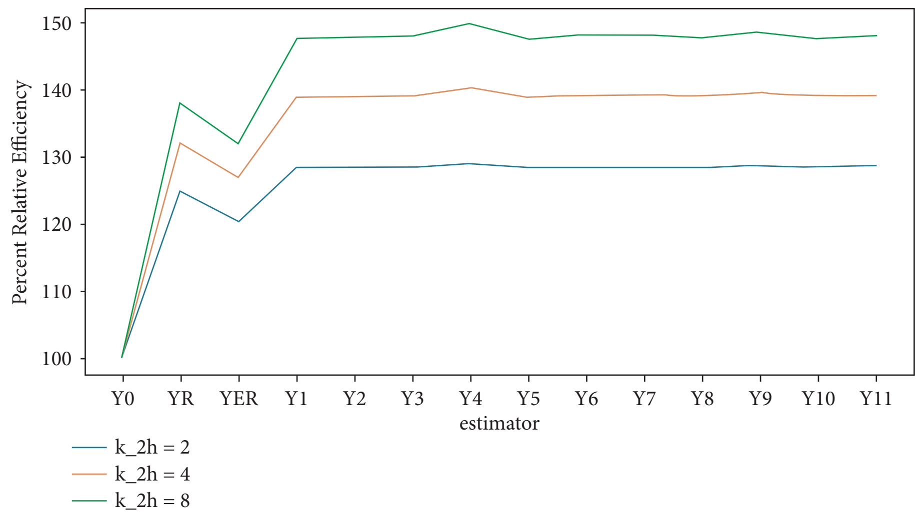

Figure 6: Plot of PREs at 20% nonresponse and 0.8 sensitivity level for population II.

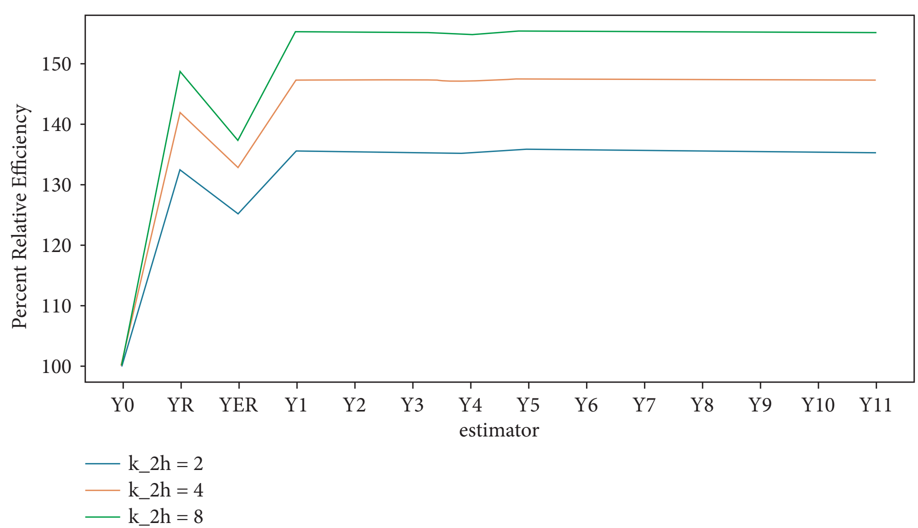

Figure 7: Plot of PREs at 30% nonresponse and 0.2 sensitivity level for population II.

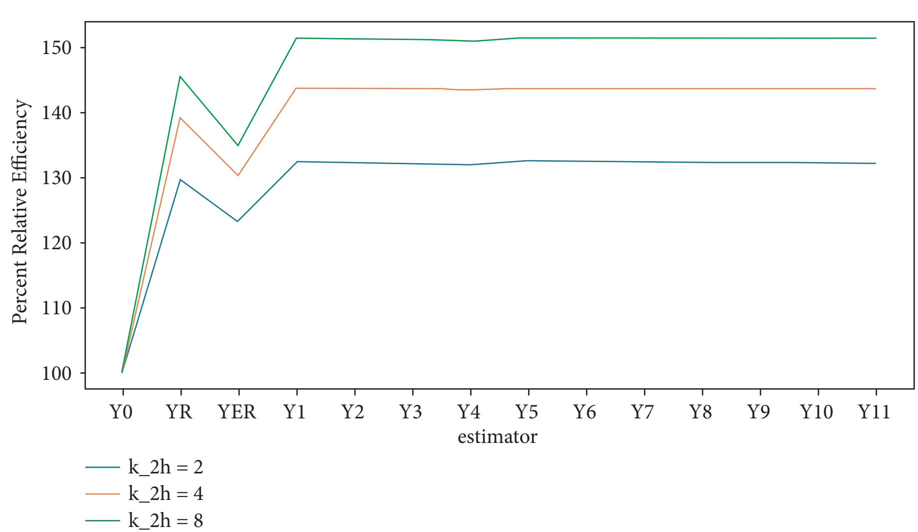

Figure 8: Plot of PREs at 30% nonresponse and 0.8 sensitivity level for population II.

8. Conclusion

Using auxiliary information, this study proposes a generalized class of exponential ratio-type estimators in the presence of nonresponse and the three-stage RRT model. The theoretical properties of bias and mean squared error of the proposed estimators are investigated up to the first degree of approximation. The theoretical performance of the proposed mean estimators is investigated. The applicability of the proposed mean estimators in practice is demonstrated using two different datasets.

According to the numerical analysis, the efficiency of the mean estimators increases as inverse sampling rates and nonresponse rates increase. Furthermore, as the sensitivity level of the survey question increases, the values for PREs decrease. The most important result of the study is that the proposed estimators outperform existing mean estimators. In the future, the proposed method could be used to estimate other population parameters of sensitive variables, like variance and distribution function in stratified two-phase sampling.

Data Availability

The data are included within the study for finding the results.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

[1] S. L. Warner, "Randomized response: a survey technique for eliminating evasive answer bias," *Journal of the American Statistical Association*, vol. 60, no. 309, pp. 63–69, 1965.

[2] K. H. Pollock and Y. Bek, "A comparison of three randomized response models for quantitative data," *Journal of the American Statistical Association*, vol. 71, no. 356, pp. 884–886, 1976.

[3] A. Chaudhuri and R. Mukherjee, *Randomized Response: Theory and Techniques*, Marcel Dekker, New York, NY, USA, 1988.

[4] S. Gupta and J. Shabbir, "Sensitivity estimation for personal interview survey questions," *Statistica*, vol. 64, no. 4, pp. 643–653, 2004.

[5] S. Gupta, J. Shabbir, and S. Sehra, "Mean and sensitivity estimation in optional randomized response models," *Journal of Statistical Planning and Inference*, vol. 140, no. 10, pp. 2870–2874, 2010.

[6] S. Mehta, B. K. Dass, J. Shabbir, and S. Gupta, "A three-stage optional randomized response model," *Journal of Statistical Theory and Practice*, vol. 6, no. 3, pp. 412–427, 2012.

[7] T. Neeraj and M. Prachi, "Additive randomized response model with known sensitivity level," *International Journal of Computational and Theoretical Statistics*, vol. 4, no. 2, 2017.

[8] R. Sousa, J. Shabbir, P. C. Real, and S. Gupta, "Ratio estimation of the mean of a sensitive variable in the presence of auxiliary information," *Journal of Statistical Theory and Practice*, vol. 4, no. 3, pp. 495–507, 2010.

[9] S. Gupta, T. Zatezalo, and J. Shabbir, "A Generalized mixture estimator of the mean of a sensitive variable in the presence of non-sensitive auxiliary information," *Statistical and Applications*, vol. 15, no. 1&2, pp. 27–36, 2017.

[10] N. Mushtaq, M. Noor-ul-Amin, and M. Hanif, "Estimation of population mean of a sensitive variable in stratified two-phase sampling," *Pakistan Journal of Statistics*, vol. 32, no. 1, pp. 393–404, 2016.

[11] N. Mushtaq, N. U. Amin, and M. Hanif, "A family of estimators of a sensitive variable using auxiliary information in stratified random sampling," *Pakistan Journal of Statistics and Operation Research*, vol. 13, no. 1, pp. 141–155, 2017.

[12] U. Shahzad, M. Hanif, N. Koyuncu, and A. V. G Luengo, "A new family of estimators for mean estimation alongside the sensitivity issue," *Journal of Reliability and Statistical Studies*, vol. 10, no. 2, pp. 43–63, 2017.

[13] M. H. Hansen and W. N. Hurwitz, "The problem of non-response in sample surveys," *Journal of the American Statistical Association*, vol. 41, no. 236, pp. 517–529, 1946.

[14] M. Khalid, G. N. Singh, and J. M. Kim, "Some imputation method to deal with the problems of missing data in two-occasion successive sampling," *Communications in Statistics-Simulation and Computation*, vol. 50, 2021.

[15] M. Khalid and G. N. Singh, "Some imputation methods to deal with issue of missing data problems due to random non–response in two–occasion successive sampling," *Communications in Statistics-Simulation and Computation*, 2020.

[16] M. Khalid and G. N. Singh, "General procedures of population mean in two-occasion successive sampling under random non-response," *The Proceedings of the National Academy of Sciences, India, Section A: Physical Sciences*, 2020.

[17] U. Shahzad, M. Hanif, N. Koyuncu, and A. V. G. Luengo, "A family of ratio estimators in stratified random sampling utilizing auxiliary attribute alongside the nonresponse," *Journal of Statistical Theory and Applications*, vol. 18, no. 1, pp. 12–25, 2019.

[18] E. Zahid, J. Shabbir, S. Gupta, R. Onyango, and S. Saeed, "A generalized class of estimators for sensitive variable in the presence of measurement error and non-response," *PLoS One*, vol. 17, no. 1, Article ID e0261561, 2022.

[19] N. Naeem and J. Shabbir, "Use of a scrambled response on two occasion's successive sampling under nonresponse," *Hacettepe Journal of Mathematics and Statistics*, vol. 47, no. 3, pp. 675–684, 2018.

[20] Q. Zhang, S. Khalil, and S. Gupta, "Mean estimation in the simultaneous presence of measurement errors and non-response using optional RRT models under stratified sampling," *Journal of Statistical Computation and Simulation*, vol. 91, no. 17, pp. 3492–3504, 2021.

[21] M. Abid, N. Abbas, H. Z. Nazir, T. Nawaz, and R. A. K. Sherwani, "Improved ratio estimators for the population mean using non-conventional measures of dispersion," *Pakistan Journal of Statistics and Operation Research*, vol. 12, no. 2, pp. 353–367, 2016.

[22] I. M. Almanjahie, A. Ibrahim Al-Omari, J. Ekpenyong, and M. Subzar, "Generalized class of mean estimators with known measures for outliers treatment," *Computer Systems Science and Engineering*, vol. 38, no. 1, pp. 1–15, 2021.

[23] K. Y. Subhash, S. S. Mishra, K. S. Alok, K. Shakti, and S. Ram, "Use of non-conventional measures of dispersion for improved estimation of population," *American Journal of Operational Research*, vol. 6, no. 3, pp. 69–75, 2016.

[24] U. Shahzad, N. H. Al-Noor, M. Hanif, and I. Sajjad, "An Exponential Family of Median Based Estimators for Mean Estimation with Simple Random Sampling Scheme," *Communication In Statistics-Theory And Methods*, vol. 50, 2021.

[25] U. Shahzad, N. H. Al-Noor, N. Afshan, D. A. Alilah, M. Hanif, and M. M. Anas, "Minimum Covariance Determinant-Based Quantile Robust Regression-type Estimators for Mean Parameter," *Mathematical Problems In Engineering*, vol. 2021, Article ID 5255839, , 2021.

[26] J. L. Hodges and E. L. Lehmann, "Estimates of location based on rank tests," *The Annals of Mathematical Statistics*, vol. 34, no. 2, pp. 598–611, 1963.

[27] J. W. Tukey, *Exploratory Data Analysis*, Addison–Wesley Publishing Co., Reading, MA, USA, 1970.

[28] B. Rosner, *Fundamentals of Biostatistics*, Duxbury Press, London, UK, 2015.