Quest Journals
Journal of Research in Applied Mathematics
Volume 7 ~ Issue 9 (2021) pp: 56-64
ISSN(Online) : 2394-0743 ISSN (Print): 2394-0735
www.questjournals.org

Research Paper

Estimation of Population Mean in Poststratified Sampling Schemein the Presence of Nonresponse

[1]Izunobi Chinyeaka Hostensia, Onyeka Aloysius Chijioke,Iwueze Iheanyi Sylvester and Ogbonna Chukwudi Justin
Department of Statistics, School of Physical Sciences, Federal University of Technology, Owerri, Nigeria
Corresponding author: Izunobi, C.H. [hostensia.izunobi@futo.edu.ng]

ABSTRACT: *Nonresponse is a major challenge in surveys. The present study focuses on the development of adequate sampling strategies for the handling of the problem of nonresponse in poststratified sampling scheme. In developing the sampling strategies, the use of auxiliary information is employed, where there is nonresponse on both the study and auxiliary variables. Based on a subsampling design of the nonresponse group, three (3) separate-type estimators of the population mean of the study variable are proposed. These are the customary- type sample mean estimator, the difference-type estimator, and a class of ratio/product-type estimators. Properties of the proposed estimators are obtained both for an achieved sample configuration (conditional argument) and for repeated samples of fixed size n (unconditional argument). Efficiency conditions under which the proposed estimators would perform better than the usual poststratified sample mean estimator are obtained, as well as the best (optimum) estimators among the proposed estimators. The theoretical results were verified using numerical illustrations.*

KEYWORDS: *Nonresponse, Poststratification, Auxiliary information, Sampling strategies.*

Received 09 September, 2021; Revised: 22 September, 2021; Accepted 24 September, 2021 © The author(s) 2021. Published with open access at www.questjournals.org

I. INTRODUCTION

One of the earliest works on the handling of the problem of nonresponse in sample surveys is attributed to [1] Hansen and Hurwitz (1946). They considered the method of subsampling the nonresponse group in order to adjust for the nonresponse in a mail survey. A number of authors followed [1] in discussing ways of handling nonresponse when estimating the population mean and other parameters of interest under simple random sam- pling scheme. These include authors like [2], [3] and [4]. The work carried out by [5] involved utilization of auxiliary information by considering a ratio-type estimator for handling nonresponse in simple random sampling scheme, while [6] suggested a regression-type estimator. [7] and [8] also utilized auxiliary information for esti- mation in the presence of nonresponse under the simple random sampling scheme. [9] suggested an exponential ratio-type class of estimators, while [10], [11], [12], [13], [14] and [15] considered general classes of estimators under simple random sampling in the presence of nonresponse. [16] considered estimation of population mean under the systematic sampling scheme, while [17] estimated domain mean in the presence of nonresponse under the stratified sampling scheme. [18] and [19] considered estimation of population mean and total respectively under nonresponse in stratified random sampling scheme. From the literature, the problem of nonresponse has been considered under different sampling schemes like simple random sampling, systematic and stratified ran- dom sampling. However, literature did not reveal the availability of any sampling strategies for the handling of nonresponse in poststratified sampling scheme, hence the vacuum the present study intends to fill.

II. PROPOSED SAMPLING STRATEGIES

To formulate sampling strategies for the handling of the problem of nonresponse under the poststratified sampling scheme, consider a population, $\Omega = \{U_1, U_2, \ldots, U_N\}$, consisting of N units. Let the population be divided into subpopulations, having sizes, N_h, $h = 1, 2, \ldots, L$, such that $\sum_{h=1}^{L} N_h = N$. Let a sample of n uints be drawn from the N units in the population, using the simple random sampling without replacement (SRSWOR) method, for estimating the population mean of a study variable under the poststratified sampling scheme. We assume that n'_1 units responded out of the selected n units in the sample, while n'_2 units did not respond. The n'_1 units that responded were allocated to their various strata where n_{1h} units fall into stratum h such that $\sum_{h=1}^{L} n_{1h} = n'_1$ units. To reduce the effect of the nonresponse, a subsample of m units is selected from the nonresponse sample group of n'_2 units, and for these m subsampled units, extra efforts are made to obtain information on the missing study and auxiliary variables as well as the stratification variable. The m subsampled units were then allocated to their various strata with m_h units falling into stratum h such that $\sum_{h=1}^{L} m_h = m$ units. We assume that both n'_1 and m are large enough such that $P(n_{1h} = 0) = P(m_h = 0) = 0 \ \forall \ h$.

The above sampling design is aimed at reducing the effect of nonresponse when there are missing values on both the study and auxiliary variables when using the postratified sampling scheme. Under the proposed sampling design, the available sample data on the study (y) and auxiliary (x) variables consist of the observed values of y and x from the response sample group of n_{1h} units in stratum h, supplemented with the observed values of y and x from the nonresponse subsample group of m_h units in stratum h. Consequently, the associated sample means, $\bar{y}_{1h}(\bar{x}_{1h})$ and $\bar{y}_{2mh}(\bar{x}_{2mh})$, form the basis for constructing estimators of the population mean $(\bar{Y})$ of the study variable, y, when there is nonresponse on both the study and auxiliary variables in poststratified sampling scheme. For the h^{th} stratum, we give an expression of the sample mean of the study variable, y, as a linear function of the response and nonresponse subsample means, given by:

$$\bar{y}_h^* = \left(\frac{n_{1h}}{n_h}\right)\bar{y}_{1h} + \left(\frac{n_{2h}}{n_h}\right)\bar{y}_{2mh} \tag{1}$$

and the corresponding composite sample mean of the auxiliary variable, x, given by:

$$\bar{x}_h^* = \left(\frac{n_{1h}}{n_h}\right)\bar{x}_{1h} + \left(\frac{n_{2h}}{n_h}\right)\bar{x}_{2mh} \tag{2}$$

where $\bar{y}_{1h} = \frac{1}{n_{1h}}\sum_{i=1}^{n_{1h}} y_{hi}$ is the sample mean of y in stratum h, based on the response group sample size, n_{1h}

$\bar{y}_{2mh} = \frac{1}{m_h}\sum_{i=1}^{m_h} y_{hi}$ is the sample mean of y in stratum h, based on the nonresponse subsample group of size, m_h

$\bar{x}_{1h} = \frac{1}{n_{1h}}\sum_{i=1}^{n_{1h}} x_{hi}$ is the sample mean of x in stratum h, based on the response group sample size, n_{1h}

$\bar{x}_{2mh} = \frac{1}{m_h}\sum_{i=1}^{m_h} x_{hi}$ is the sample mean of x in stratum h, based on the nonresponse subsample group of size, m_h

Based on the above sampling design and using the nonresponse sample means (1) and (2), we propose the following three estimators of the population mean $(\bar{Y})$ of the study variable, y, when there is nonresponse on both the study and auxiliary variables under the poststratified sampling scheme.

1. The customary nonresponse sample mean estimator given by

$$\bar{y}_{ns1} = \sum_{h=1}^{L} W_h \bar{y}_h^* \tag{3}$$

2. A difference-type estimator given by

$$\bar{y}_{nd1} = \sum_{h=1}^{L} W_h [\bar{y}_h^* - k_h(\bar{x}_h^* - \bar{X}_h)] \tag{4}$$

3. A general class of ratio/product type estimators that makes use of known values of some population parameters of the auxiliary variable in the h^{th} stratum, given by

$$\bar{y}_{nc1} = \sum_{h=1}^{L} W_h \bar{y}_h^* \Big(\frac{a_h \bar{X}_h + b_h}{a_h \bar{x}_h^* + b_h}\Big)^{\alpha_h} \tag{5}$$

where $k_h(\alpha_h)$ can be any real numbers and $a_h(b_h)$ can also be any real numbers or in particular, some known values of population parameters of an auxiliary variable, x, in the h^{th} stratum, like variance, coefficient of variation, standard deviation, skewness, kurtosis, etc. We note that the estimator (3) does not make use of any auxiliary information, while the estimators (4) and (5) utilize information on one auxiliary variable. Particularly, the proposed general class of estimators (5) utilizes information on known values of population parameters of an auxiliary variable, and it is a simplified version (ratio/product type estimators) of the class of estimators suggested by [20] under the simple random sampling scheme.

III. PROPERTIES OF THE PROPOSED ESTIMATORS

We obtained the properties of the above proposed three estimators both for a given an achieved sample configuration, say $\underline{n} = (n_1, n_2, \ldots, n_L)$, (the conditional argument) and for repeated samples of fixed size, n (the unconditional argument). The results are shown in Theorem 3.1.

Therorem 3.1: The variances and mean square errors of the proposed estimators are obtained under the conditional argument as follows:

$$V_c(\bar{y}_{ns1}) = \sum_{h=1}^{L} W_h^2 \Big(\frac{1}{n_h} - \frac{1}{N_h}\Big) S_{yh}^2 + \sum_{h=1}^{L} W_h^2 \Big(\frac{g_h - 1}{n_h}\Big) W_{2h} S_{2yh}^2 \tag{6}$$

$$\begin{aligned} V_c(\bar{y}_{nd1}) = {} & \sum_{h=1}^{L} W_h^2 \Big(\frac{1}{n_h} - \frac{1}{N_h}\Big)(S_{yh}^2 + k_h^2 S_{xh}^2 - 2k_h S_{yxh}) \\ & + \sum_{h=1}^{L} W_h^2 \Big(\frac{g_h - 1}{n_h}\Big) W_{2h} (S_{2yh}^2 + k_h^2 S_{2xh}^2 - 2k_h S_{2yxh}) \end{aligned} \tag{7}$$

$$\begin{aligned} MSE_c(\bar{y}_{nc1}) = {} & \sum_{h=1}^{L} W_h^2 \Big(\frac{1}{n_h} - \frac{1}{N_h}\Big)(S_{yh}^2 + \alpha_h^2 \lambda_h^2 R_h^2 S_{xh}^2 - 2\alpha_h \lambda_h R_h S_{yxh}) \\ & + \sum_{h=1}^{L} W_h^2 \Big(\frac{g_h - 1}{n_h}\Big) W_{2h} (S_{2yh}^2 + \alpha_h^2 \lambda_h^2 R_h^2 S_{2xh}^2 - 2\alpha_h \lambda_h R_h S_{2yxh}) \end{aligned} \tag{8}$$

Furthermore, the variances and mean square errors of the estimators are obtained under the unconditional argument, and up to first order of approximations as follows:

$$V(\bar{y}_{ns1}) = \Big(\frac{1-f}{n}\Big) \sum_{h=1}^{L} W_h S_{yh}^2 + \Big(\frac{1}{n}\Big) \sum_{h=1}^{L} W_h (g_h - 1) W_{2h} S_{2yh}^2 \tag{9}$$

$$\begin{aligned} V(\bar{y}_{nd1}) = {} & \Big(\frac{1-f}{n}\Big) \sum_{h=1}^{L} W_h (S_{yh}^2 + k_h^2 S_{xh}^2 - 2k_h S_{yxh}) \\ & + \Big(\frac{1}{n}\Big) \sum_{h=1}^{L} W_h (g_h - 1) W_{2h} (S_{2yh}^2 + k_h^2 S_{2xh}^2 - 2k_h S_{2yxh}) \end{aligned} \tag{10}$$

$$MSE(\bar{y}_{nc1}) = \Big(\frac{1-f}{n}\Big)\sum_{h=1}^{L} W_h(S_{yh}^2 + \alpha_h^2\lambda_h^2R_h^2S_{xh}^2 - 2\alpha_h\lambda_hR_hS_{yxh}) + \Big(\frac{1}{n}\Big)\sum_{h=1}^{L} W_h(g_h-1)W_{2h}(S_{2yh}^2 + \alpha_h^2\lambda_h^2R_h^2S_{2xh}^2 - 2\alpha_h\lambda_hR_hS_{2yxh}) \tag{11}$$

where

$$\lambda_h = \Big(\frac{a_h\bar{X}_h}{a_h\bar{X}_h + b_h}\Big) \tag{12}$$

Proof: We first obtain the properties of the proposed customary sample mean estimator, $\bar{y}_{ns1}$. Let E_1, V_1 and C_1 respectively denote expectation, variance and covariance over the initial or first phase sampling of n units out of N units in the population, while E_2, V_2 and C_2 respectively denote expectation, variance and covariance over the subsample or second phase sampling of m units out of the n_2' nonresponse units. Then from (1) we have:

$$E_2(\bar{y}_h^*) = \Big(\frac{n_{1h}}{n_h}\Big)E_2(\bar{y}_{1h}) + \Big(\frac{n_{2h}}{n_h}\Big)E_2(\bar{y}_{2mh}) = \Big(\frac{n_{1h}}{n_h}\Big)(\bar{y}_{1h}) + \Big(\frac{n_{2h}}{n_h}\Big)(\bar{y}_{2h}) = \bar{y}_h \tag{13}$$

and

$$V_2(\bar{y}_h^*) = \Big(\frac{n_{1h}}{n_h}\Big)^2V_2(\bar{y}_{1h}) + \Big(\frac{n_{2h}}{n_h}\Big)^2V_2(\bar{y}_{2mh}) + 2\Big(\frac{n_{1h}}{n_h}\Big)\Big(\frac{n_{2h}}{n_h}\Big)C_2(\bar{y}_{1h}, \bar{y}_{2mh}) \tag{14}$$

Considering the fact that $V_2(\bar{y}_{1h}) = C_2(\bar{y}_{1h}, \bar{y}_{2mh}) = 0$ and applying the theory of simple random sampling without replacement (SRSWOR) method, we write (14) as

$$V_2(\bar{y}_h^*) = \Big(\frac{n_{2h}}{n_h}\Big)^2V_2(\bar{y}_{2mh}) = \Big(\frac{n_{2h}}{n_h}\Big)^2\Big(\frac{1}{m_h} - \frac{1}{n_{2h}}\Big)s_{2yh}^2 = \Big(\frac{g_h-1}{n_h}\Big)\Big(\frac{n_{2h}}{n_h}\Big)s_{2yh}^2 \tag{15}$$

To obtain the conditional expectation of $\bar{y}_h^*$, we apply the theories of double phase sampling and SRSWOR methods and use (13), to obtain

$$E_c(\bar{y}_h^*) = E_1E_2(\bar{y}_h^*) = E_1(\bar{y}_h) = \bar{Y}_h \tag{16}$$

Using (3) and (16), the conditional expectation of the proposed customary nonresponse sample mean estimator, $\bar{y}_{ns1}$, is obtained as

$$E_c(\bar{y}_{ns1}) = \sum_{h=1}^{L} W_hE_c(\bar{y}_h^*) = \sum_{h=1}^{L} W_h\bar{Y}_h = \bar{Y} \tag{17}$$

Hence, $\bar{y}_{ns1}$ is unbiased for $\bar{Y}$ under the conditional argument. To obtain the conditional variance of $\bar{y}_{ns1}$, we first obtain the conditional variance of $\bar{y}_h^*$ in the h^{th} stratum. Applying the theory of double sampling, the conditional variance of $\bar{y}_h^*$ can be written as:

$$V_c(\bar{y}_h^*) = V_1E_2(\bar{y}_h^*) + E_1V_2(\bar{y}_h^*) \tag{18}$$

Substituting (13) and (15) into (18) and applying results from SRSWOR gives:

$$V_c(\bar{y}_h^*) = V_1(\bar{y}_h) + E_1\Big[\Big(\frac{g_h-1}{n_h}\Big)\Big(\frac{n_{2h}}{n_h}\Big)s_{2yh}^2\Big] = V_1(\bar{y}_h) + \Big(\frac{g_h-1}{n_h}\Big)E_1\Big(\frac{n_{2h}}{n_h}\Big)E_1(s_{2yh}^2) \tag{19}$$

or

$$V_c(\bar{y}_h^*) = \Big(\frac{1}{n_h} - \frac{1}{N_h}\Big)S_{yh}^2 + \Big(\frac{g_h-1}{n_h}\Big)W_{2h}S_{2yh}^2 \tag{20}$$

From (3), we write the conditional variance of $\bar{y}_{ns1}$ as

$$V_c(\bar{y}_{ns1}) = V_c\Big(\sum_{h=1}^{L} W_h\bar{y}_h^*\Big) = \sum_{h=1}^{L} W_h^2V_c(\bar{y}_h^*) + 2\sum_{h=1}^{L}\sum_{j>h}^{L} W_hW_jC_c(\bar{y}_h^*, \bar{y}_j^*) = \sum_{h=1}^{L} W_h^2V_c(\bar{y}_h^*) \tag{21}$$

since the covariance term between two different strata means vanishes. Substituting (20) in (21) gives the conditional variance of the proposed customary nonresponse sample mean estimator, $\bar{y}_{ns1}$, as

$$V_c(\bar{y}_{ns1}) = \sum_{h=1}^{L} W_h^2\Big(\frac{1}{n_h} - \frac{1}{N_h}\Big)S_{yh}^2 + \sum_{h=1}^{L} W_h^2\Big(\frac{g_h-1}{n_h}\Big)W_{2h}S_{2yh}^2 \tag{22}$$

as stated in the theorem. This proves the aspect of the first part of the theorem that bothers on the conditional variance of the proposed customary nonresponse sample mean estimator, $\bar{y}_{ns1}$, of the population mean ($\bar{Y}$) of the study variable, y. To prove the corresponding second part of the theorem, which bothers on the unconditional variance of the proposed nonresponse estimator, $\bar{y}_{ns1}$, we consider repeated samples of fixed size of n units selected under the poststratified sampling scheme in which the h^{th} stratum sample size, n_h, is a random variable. Following [21], the stratum sample size, n_h, has a hypergeometric distribution with

$$E\Big(\frac{1}{n_h}\Big) = \frac{1}{nW_h} \tag{23}$$

$$E(\bar{y}_{ns1}) = EE_c(\bar{y}_{ns1}) = E(\bar{Y}) = \bar{Y} \tag{24}$$

Hence, the proposed customary nonresponse sample mean estimator, $\bar{y}_{ns1}$, is unbiased for the population mean ($\bar{Y}$) under the unconditional argument. Furthermore, taking the unconditional expectation of (22) gives the unconditional variance of $\bar{y}_{ns1}$ as:

$$V(\bar{y}_{ns1}) = EV_c(\bar{y}_{ns1}) = \sum_{h=1}^{L} W_h^2\Big[E\Big(\frac{1}{n_h}\Big) - \frac{1}{N_h}\Big]S_{yh}^2 + \sum_{h=1}^{L} W_h^2 E\Big(\frac{1}{n_h}\Big)(g_h - 1)W_{2h}S_{2yh}^2 \tag{25}$$

and using (23) to make the necessary substitutions in (25) gives the unconditional variance of the proposed sample mean estimator, $\bar{y}_{ns1}$, up to first order of approximations, as:

$$V(\bar{y}_{ns1}) = \Big(\frac{1-f}{n}\Big)\sum_{h=1}^{L} W_h S_{yh}^2 + \Big(\frac{1}{n}\Big)\sum_{h=1}^{L} W_h(g_h - 1)W_{2h}S_{2yh}^2 \tag{26}$$

as stated in the theorem. Following similar procedure, the conditional and unconditional properties of the proposed difference-type estimator, $\bar{y}_{nd1}$, and the proposed class of ratio/product estimators, $\bar{y}_{nc1}$, are obtained as stated in the theorem. And this completes the proof.

Notice that the proposed class of estimators, $\bar{y}_{nc1}$, given in (5), is capable of generating a wide range of estimators of the population mean ($\bar{Y}$) of the study variable, y, when there is nonresponse on both the study and auxiliary variables. This is achieved by making appropriate choices of the constants, α_h, a_h and b_h. Some special cases of the proposed class of estimators, $\bar{y}_{nc1}$, obtained by making suitable choices of the three constants, α_h, a_h and b_h, in the h^{th} stratum, are as shown in Table 3.1. The best estimators in the proposed class of estimators, $\bar{y}_{nc1}$, are those with the best choices of the three constants, α_h, a_h and b_h, in the h^{th} stratum. Accordingly, the best choices of the constants, α_h, a_h and b_h, are those that minimize the conditional and unconditional mean square errors of the proposed class of estimators, $\bar{y}_{nc1}$, as the case may be. The results we obtained are given in theorem 3.2.

Theorem 3.2: Under the conditional argument, the values of the constants, α_h, a_h and b_h, that minimize the conditional mean square error of the proposed class of estimators, $\bar{y}_{nc1}$, are obtained as

$$\theta'_{h1} = \frac{\alpha_h a_h \bar{X}_h R_h}{a_h\bar{X}_h + b_h} = \frac{\alpha_h a_h \bar{Y}_h}{a_h\bar{X}_h + b_h} = \frac{\big(\frac{1}{n_h} - \frac{1}{N_h}\big)S_{yxh} + \big(\frac{g_h-1}{n_h}\big)S_{2yxh}}{\big(\frac{1}{n_h} - \frac{1}{N_h}\big)S_{xh}^2 + \big(\frac{g_h-1}{n_h}\big)S_{2xh}^2} \tag{27}$$

$$\theta_{h1} = \frac{\alpha_h a_h \bar{X}_h R_h}{a_h\bar{X}_h + b_h} = \frac{\alpha_h a_h \bar{Y}_h}{a_h\bar{X}_h + b_h} = \frac{(1-f)S_{yxh} + (g_h - 1)W_{2h}S_{2yxh}}{(1-f)S_{xh}^2 + (g_h - 1)W_{2h}S_{2xh}^2} \tag{28}$$

Proof: The conditional and unconditional mean square errors of the proposed class of estimators, $\bar{y}_{nc1}$, respectively given as (8) and (11), in theorem 3.1, can be written as:

$$\begin{aligned} MSE_c(\bar{y}_{nc1}) = &\sum_{h=1}^{L} W_h^2\Big(\frac{1}{n_h} - \frac{1}{N_h}\Big)(S_{yh}^2 + \theta_h^2 S_{xh}^2 - 2\theta_h S_{yxh}) \\ &+ \sum_{h=1}^{L} W_h^2\Big(\frac{g_h - 1}{n_h}\Big)W_{2h}(S_{2yh}^2 + \theta_h^2 S_{2xh}^2 - 2\theta_h S_{2yxh}) \end{aligned} \tag{29}$$

or

$$\begin{aligned} MSE(\bar{y}_{nc1}) = &\Big(\frac{1-f}{n}\Big)\sum_{h=1}^{L} W_h(S_{yh}^2 + \theta_h^2 S_{xh}^2 - 2\theta_h S_{yxh}) \\ &+ \Big(\frac{1}{n}\Big)\sum_{h=1}^{L} W_h(g_h - 1)W_{2h}(S_{2yh}^2 + \theta_h^2 S_{2xh}^2 - 2\theta_h S_{2yxh}) \end{aligned} \tag{30}$$

where

$$\theta_h = \frac{\alpha_h a_h \bar{X}_h R_h}{a_h\bar{X}_h + b_h} = \frac{\alpha_h a_h \bar{Y}_h}{a_h\bar{X}_h + b_h} \tag{31}$$

Using the method of least squares to minimize (29) and (30) with respect to θ_h, the optimum values, (27) and (28), are obtained as given in the theorem. This completes the proof.

Table 3.1: Special Cases of Proposed Class of Estimators

SN	Estimator	α_h	a_h	b_h
1	$\bar{y}_{nc1}(1) = \sum_{h=1}^{L} W_h \bar{y}_h^* \left(\bar{X}_h / \bar{x}_h^*\right)$, Customary ratio-type estimator	1	1	0
2	$\bar{y}_{nc1}(2) = \sum_{h=1}^{L} W_h \bar{y}_h^* \left(\bar{X}_h + C_{xh}\right) / \left(\bar{x}_h^* + C_{xh}\right)$, [22] Sisodia and Dwivedi (1981) estimator	1	1	C_{xh}
3	$\bar{y}_{nc1}(3) = \sum_{h=1}^{L} W_h \bar{y}_h^* \left(\bar{X}_h + \beta_{2xh}\right) / \left(\bar{x}_h^* + \beta_{2xh}\right)$, [23] Singh and Kakran (1993) estimator I	1	1	β_{2xh}
4	$\bar{y}_{nc1}(4) = \sum_{h=1}^{L} W_h \bar{y}_h^* \left(\bar{X}_h C_{xh} + \beta_{2xh}\right) / \left(\bar{x}_h^* C_{xh} + \beta_{2xh}\right)$, [24] Upadhyaya and Singh (1999) estimator I	1	C_{xh}	β_{2xh}
5	$\bar{y}_{nc1}(5) = \sum_{h=1}^{L} W_h \bar{y}_h^* \left(\bar{X}_h \beta_{2xh} + C_{xh}\right) / \left(\bar{x}_h^* \beta_{2xh} + C_{xh}\right)$, [24] Upadhyaya and Singh (1999) estimator II	1	β_{2xh}	C_{xh}
6	$\bar{y}_{nc1}(6) = \sum_{h=1}^{L} W_h \bar{y}_h^* \left(\bar{x}_h^* / \bar{X}_h\right)$, Customary product-type estimator	−1	1	0
7	$\bar{y}_{nc1}(7) = \sum_{h=1}^{L} W_h \bar{y}_h^* \left(\bar{x}_h^* + C_{xh}\right) / \left(\bar{X}_h + C_{xh}\right)$, [25] Pandey and Dubey (1988) estimator	−1	1	C_{xh}
8	$\bar{y}_{nc1}(8) = \sum_{h=1}^{L} W_h \bar{y}_h^* \left(\bar{x}_h^* + \beta_{2xh}\right) / \left(\bar{X}_h + \beta_{2xh}\right)$, [23] Singh and Kakran (1993) estimator II	−1	1	β_{2xh}
9	$\bar{y}_{nc1}(9) = \sum_{h=1}^{L} W_h \bar{y}_h^* \left(\bar{x}_h^* C_{xh} + \beta_{2xh}\right) / \left(\bar{X}_h C_{xh} + \beta_{2xh}\right)$, [24] Upadhyaya and Singh (1999) estimator III	−1	C_{xh}	β_{2xh}
10	$\bar{y}_{nc1}(10) = \sum_{h=1}^{L} W_h \bar{y}_h^* \left(\bar{x}_h^* \beta_{2xh} + \sigma_{xh}\right) / \left(\bar{X}_h \beta_{2xh} + \sigma_{xh}\right)$, [26] Singh (2003) estimator	−1	β_{2xh}	σ_{xh}

IV. NUMERICAL ILLUSTRATION

For the numerical illustration of the theoretical results, we shall use the Australian Survey data ([27] Telford and Cunningham (1991)), with sum of skin folds (ssf) and the percentage Body fat (pcBfaT) as the study (y) and auxiliary (x) variables respectively. Here, the study and auxiliary variables are positively correlated. We have stratified the respondents with respect to their sex, where the females constitute Stratum 1 and the males as Stratum 2. We take a random sample of $n = 60$ respodents from the population of $N = 202$ respondents. The summary statistics of the data are shown in Table 4.1, using the special cases given in Table 3.1.

Table 4.1: Summary Statistics for Australian Survey data

Parameters	Stratum 1 (Females)	Stratum 2 (Males)
N_h	100	102
N_{2h}	40	42
W_h	0.495050	0.504951
W_{2h}	0.4	0.411765
$\bar{Y}_h$	17.849100	9.250882
$\bar{X}_h$	86.973000	51.422549
R_h	0.205226	0.179899
S_{yh}^2	29.734018	10.142270
S_{xh}^2	1145.860577	355.476219
S_{yxh}	178.960117	58.079227
S_{2yh}^2	12.696796	16.307434
S_{2xh}^2	517.849942	552.476986
S_{2yxh}	74.617189	92.682720
C_{xh}	0.389208	0.366650
β_{2xh}	0.775909	1.385599

Source: Australian Survey Data on sex, sport and body-size dependency of hematology in highly trained athletes by [27] Telford and Cunningham (1991)

The Percentage Relative Efficiencies (PRE) of the proposed estimators over the usual poststratified sample mean estimator, $(\bar{y}_{ps})$, shall be obtained. The results are shown in Table 4.2. Table 4.2 shows the percentage relative efficiencies of some proposed estimators over the usual poststratified sample mean estimator $(\bar{y}_{ps})$, when there is nonresponse on both the study and auxiliary variables. Because of the positive correlation between the study and auxiliary variables, the proposed ratio-type estimators, as expected, performed better than the proposed product-type estimators, in terms of having higher percentage relative efficiencies. Also, the optimum estimators in the proposed class of estimators, $(\bar{y}_{nc1})$, as expected, gave the highest values of the percentage relative efficiencies, especially with increase in the subsample size, m_h. Generally, we observed from Table 4.2 that the efficiencies of the proposed estimators, in terms of having smaller mean square errors and higher pecentage relative efficiencies, increase with an increase in the subsampling fraction $(1/g_h)$. With subsampling fractions of 0.10 (10%) and above, some of the proposed ratio-type estimators showed high percentage relative efficiencies (above 100%) over the usual poststratified sample mean estimator, $\bar{y}_{ps}$. However, the proposed customary sample mean estimator, $\bar{y}_{ns1} = \sum_{h=1}^{L} W_h\bar{y}_h^*$, could not provide a better estimate than the usual poststratified sample mean estimator, $(\bar{y}_{ps})$, even with a subsampling fraction of 0.90 (90%). This justifies the recourse to the use of auxiliary information in an attempt to formulate estimators/sampling strategies with increased efficiency over the usual poststratified sample mean estimator, $(\bar{y}_{ps})$. Among the ten (10) considered special cases of the proposed class of estimators, $(\bar{y}_{nc1})$, the estimator that has the highest efficiency is the adopted Upadhyaya-Singh (1999) estimator I, $\bar{y}_{nc1}(4) = \sum_{h=1}^{L} W_h\bar{y}_h^*\left(\bar{X}_hC_{xh} + \beta_{2xh}\right)/\left(\bar{x}_h^*C_{xh} + \beta_{2xh}\right)$, followed by the adopted Singh-Kakran (1993) estimator, $\bar{y}_{nc1}(3) = \sum_{h=1}^{L} W_h\bar{y}_h^*\left(\bar{X}_h + \beta_{2xh}\right)/\left(\bar{x}_h^* + \beta_{2xh}\right)$, while the estimator with the least efficiency among the ten (10) special cases under consideration is the customary product-type estimator, $\bar{y}_{nc1}(6) = \sum_{h=1}^{L} W_h\bar{y}_h^*\left(\bar{x}_h^*/\bar{X}_h\right)$, when there is nonresponse on both the study and auxiliary variables. We also observe from Table 4.2 that the efficiency of the proposed difference-type estimator, $\bar{y}_{nd1} = \sum_{h=1}^{L} W_h[\bar{y}_h^* - k_h(\bar{x}_h^* - \bar{X}_h)]$, $k_h = S_{yxh}/S_{xh}^2\ \forall\, h$, comes very close to the efficiency of the optimum estimators in the proposed class of estimators, $(\bar{y}_{nc1})$. This is because our choice of the constant, $k_h = S_{yxh}/S_{xh}^2\ \forall\, h$, is very close to its optimum value that minimizes the variance of the difference-type estimator, $\bar{y}_{nd1}$.

Table 4.2: Percentage Relative Efficiency (PRE) of Some Proposed Estimators over the Usual Poststratified Mean Estimator $(\bar{y}_{ps})$ when there is nonresponse on both the study and auxiliary variables for positively correlated study and auxiliary variables (Dataset-1)

Estimators	**Subsampling Fraction** $(1/g_h)$; where $g_h = n_{2h}/m_h > 1$ is the subsampling factor								
	0.10	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**
$\bar{y}_{ps} = \sum_{h=1}^{L} W_h\bar{y}_h$	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
$\bar{y}_{ns1} = \sum_{h=1}^{L} W_h\bar{y}_h^*$	20.79	37.13	50.31	61.16	70.26	77.99	84.64	90.43	95.51
$\bar{y}_{nd1} = \sum_{h=1}^{L} W_h[\bar{y}_h^* - k_h(\bar{x}_h^* - \bar{X}_h)]$, $k_h = S_{yxh}/S_{xh}^2\ \forall\, h$	236.62	450.87	645.78	823.86	987.18	1137.53	1276.37	1404.99	1524.47
$\bar{y}_{nc1}(1) = \sum_{h=1}^{L} W_h\bar{y}_h^*\left(\bar{X}_h/\bar{x}_h^*\right)$	134.95	248.22	344.63	427.69	500.00	563.51	619.74	669.88	714.85
$\bar{y}_{nc1}(2) = \sum_{h=1}^{L} W_h\bar{y}_h^*\left(\bar{X}_h + C_{xh}\right)/\left(\bar{x}_h^* + C_{xh}\right)$	137.25	252.69	351.13	436.07	510.11	575.22	632.92	684.41	730.65
$\bar{y}_{nc1}(3) = \sum_{h=1}^{L} W_h\bar{y}_h^*\left(\bar{X}_h + \beta_{2xh}\right)/\left(\bar{x}_h^* + \beta_{2xh}\right)$	140.16	258.26	359.15	446.31	522.39	589.36	648.77	701.83	749.50
$\bar{y}_{nc1}(4) = \sum_{h=1}^{L} W_h\bar{y}_h^*\left(\bar{X}_hC_{xh} + \beta_{2xh}\right)/\left(\bar{x}_h^*C_{xh} + \beta_{2xh}\right)$	147.13	272.10	379.58	472.98	554.92	627.37	691.89	749.72	801.85
$\bar{y}_{nc1}(5) = \sum_{h=1}^{L} W_h\bar{y}_h^*\left(\bar{X}_h\beta_{2xh} + C_{xh}\right)/\left(\bar{x}_h^*\beta_{2xh} + C_{xh}\right)$	137.66	253.52	352.39	437.74	512.17	577.66	635.71	687.54	734.08
$\bar{y}_{nc1}(6) = \sum_{h=1}^{L} W_h\bar{y}_h^*\left(\bar{x}_h^*/\bar{X}_h\right)$	4.50	7.94	10.67	12.89	14.72	16.26	17.58	18.71	19.70
$\bar{y}_{nc1}(7) = \sum_{h=1}^{L} W_h\bar{y}_h^*\left(\bar{x}_h^* + C_{xh}\right)/\left(\bar{X}_h + C_{xh}\right)$	4.52	7.99	10.74	12.96	14.81	16.35	17.68	18.82	19.81
$\bar{y}_{nc1}(8) = \sum_{h=1}^{L} W_h\bar{y}_h^*\left(\bar{x}_h^* + \beta_{2xh}\right)/\left(\bar{X}_h + \beta_{2xh}\right)$	4.58	8.08	10.85	13.09	14.95	16.51	17.83	18.98	19.98
$\bar{y}_{nc1}(9) = \sum_{h=1}^{L} W_h\bar{y}_h^*\left(\bar{x}_h^*C_{xh} + \beta_{2xh}\right)/\left(\bar{X}_hC_{xh} + \beta_{2xh}\right)$	4.71	8.30	11.13	13.42	15.31	16.90	18.25	19.41	20.42
$\bar{y}_{nc1}(10) = \sum_{h=1}^{L} W_h\bar{y}_h^*\left(\bar{x}_h^*\beta_{2xh} + \sigma_{xh}\right)/\left(\bar{X}_h\beta_{2xh} + \sigma_{xh}\right)$	6.23	11.11	15.04	18.27	20.97	23.27	25.24	26.95	28.45
Optimum estimators	237.06	451.57	646.61	824.70	987.98	1138.21	1276.90	1405.33	1524.59

It can also be observed from Table 4.2 that the use of auxiliary information does not always result in an improved efficiency over the proposed customary sample mean estimator. For the positively correlated study and auxiliary variables dataset used in this study, we observed that the proposed customary sample mean estimator, $\bar{y}_{ns1} = \sum_{h=1}^{L} W_h \bar{y}_h^*$, was less efficient than the ratio-type estimators in the proposed class of estimators, $\bar{y}_{nc1}$, but more efficient than the product-type estimators, under considerations, in the proposed class of estimators. This means that the use of auxiliary information will likely result in increased efficiency over the proposed customary sample mean estimator when auxiliary information are utilized according to the relationship between the study and auxiliary variables. Table 4.2 shows that the ratio-type estimators are to be preferred when the study and auxiliary variables are positively correlated.

V. CONCLUDING REMARKS

The proposed customary sample mean estimator is less efficient than the usual poststratified sample mean estimator, $\bar{y}_{ps}$, even with subsampling fractions of up to 0.90 (90%) in the presence of nonresponse. It can therefore be concluded that there is need to consider the use of auxiliary information in the search for more efficient estimators than the usual poststratified sample mean estimator, $\bar{y}_{ps}$, in the handling of nonresponse under the poststratified sampling scheme. The use of auxiliary information is more likely to result in an increased efficiency over the usual poststratified sample mean estimator, $\bar{y}_{ps}$, when auxiliary information are utilized according to the relationship between the study and auxiliary variables. The best estimators or the most efficient estimators among the proposed class of estimators are found to be more efficient than the usual poststratified sample mean estimator, $\bar{y}_{ps}$, as well as more efficient than other ratio-type/product-type estimators in the proposed class of estimators. For the positively correlated study and auxiliary variables used for numerial illustration, the proposed estimator with the highest efficiency among the ten (10) special cases under consideration was the adopted Upadhyaya-Singh (1999) estimator I, while the estimator with the least efficiency among the ten (10) special cases under consideration was the customary product-type estimator. On the basis of the empirical results obtained in the study, we recommend the use of the proposed adopted (ratio-type) Upadhyaya-Singh (1999) estimator I, for positively correlated study and auxiliary variables. The use of the proposed difference-type estimator is recommended only when it is possible to choose the value of the constanst, k_h, to be very close to the optimum value that minimizes the mean square error of the proposed class of estimators.

REFERENCES

[1]. Hansen, M. H. & Hurwitz, W. N. (1946). The problem of non-response in sample surveys. Journal of the American Statistical Association, 41(2), 517–529.

[2]. Kish, L. & Lansing, J. B. .(1954). Response errors in estimating the value of homes. Journal of the American Statistical Association, 49(267), 520-538, DOI: 10.1080/01621459.1954.10483519.

[3]. Little, R. J. A. (1982). Models for non response in sample surveys. J. Amer. Statist. Assoc., 77, 237–250.

[4]. Khare, B. B. & Srivastava, S. (1997). Transformed ratio type esimators for the popuiation mean in the presence of nonresponse. Communications in Statistics–Theory and Methods, 26(7), 1719–1791.

[5]. Rao, P. S. R. S. (1986). Ratio estimation with sub-sampling the non-respondents. Survey MethodoIogy, 12(2), 217–230.

[6]. Rao, P. S. R. S. (1990). Regression estimators with sub sampling of non-respondents. In: Data Quality Control. Theory and Pragmatics, (Gunar E. Liepins and V.R.R. Uppuluri, eds.), Marcel Dekker, New York, 191-208.

[7]. Khare, B. B. & Srivastava, S. (1993). Estimation of population mean using auxiliary character in presence of non-response. Nat. Acad. Sci. Lett. India, 16, 111–114.

[8]. Tripathi, T. P., Das, A. K. & Khare, B. B. (1994). Use of auxiliary information in population mean in presence of non response. Assam Statist. Rev., 20(1), 73–83.

[9]. Grover, L. K. & Kaur, P. (2014). Exponential ratio type estimators of population mean under non- response. Open Journal of Statistics, 4, 97–110.

[10]. Kumar, S. (2014). Estimation of the ratio, product and mean using multiauxiliary variables in the presence of non-response. Chilean Journal of Statistics, 5(1), 49–72.

[11]. Kumar, N., & Patel, K. M. (2015). Estimation of ratio of two population means using regression estimators in presence of non-response. Research Journal of Mathematical and Statistical Sciences,3(1), 1–5.

[12]. Sinha, R. R. (2016). Families of estimators for ratio and product of study characters using mean and proportion of auxiliary character in presence of non-response. International Journal of Accountingand Economics Studies, 4(2), 142–147.

[13]. Rachokarn, T. & Lawson, N. (2017). An efficient general family of estimators for population mean in the presence of non-response. Journal of Mathematical and Fundamental Sciences, 49(3), 283–293.

[14]. Koyuncu, N., Shahzad, U , Hanif, M. & Luengo, A. V. G. (2017). Estimation of population variance using quartiles in absence and presence of non-response. Gazi university Journal of science, 30(2),205–218.

[15]. Lawson, N., Rachokarn, T., & Charurotkeerati, T. (2018). An efficient general family of estimators for population means in sampling with non-response. Matter: International Journal of Science andTechnology, 4(2), 1–11.

[16]. Shahzad, U., Hanif, M., Koyuncu, N. & Luengo, A. V. G. (2017). Estimation of population mean un- der systematic random in the absence and presence of nonresponse. Journal of Statistician: Statistics and Acturian Sciences, IDIA, 10(1), 23–39.

[17]. Clement, E. P., Udofia, G. A. & Enang, E. I. (2014). Estimation for domains in stratified sam- pling design in the presence of nonresponse. American Journal of Mathematics and Statistics, 4(2),65–71.

[18]. Onyeka, A. C., Ogbumuo, D. T. & Izunobi, C. H. (2019). Estimation of population mean in strati- fied random sampling when using auxiliary information in the presence of non-response. Far EastJournal of Theoretical Statistics, 55(2), 151–167.

[19]. Oyoo, D., Manene, M., Ouma, C. & Muhua, G. (2019). Modified ratio estimator of finite population total in stratified random sampling under nonresponse. Mathematical Theory and Modeling, 9(7),doi:10.7176/MTM PP74-88.

[20]. Khoshnevisan, M, Singh, R, Chauhan, P, Sawan, N., & Smarandache, F (2007). A general family of estimators for estimating population

mean using known value of some population parameter(s), Far East Journal of Theoretical Statistics, 22, 181–191.

[21]. Stephan, F. (1945). The expected value and variance of the reciprocal and other negative powers of a positive Bernoulli variate. Ann. Math. Stat., 16, 50-61.

[22]. Sisodia, B. & Dwivedi, V. (1981). Modified ratio estimator using coefficient of variation of auxiliary variable. Journal-Indian Society of Agricultural Statistics.

[23]. Singh, H.P., & Kakran, M.S. (1993). A modified ratio estimator using known coefficient of kurtosis of an auxiliary character. Revised version submitted to Journal of Indian Society of Agricultural Statistics, New Delhi, India.

[24]. Upadhyaya, L. N. & Singh, H. P. (1999). Use of transformed auxiliary variable in estimating the finite population mean. Biometrical Journal: Journal of Mathematical Methods in Biosciences, 41(5), 627–636.

[25]. Pandey, B. & Dubey, V. (1988). Modified product estimator using coefficient of variation of auxiliary variate. Assam Statistical Rev, 2(2), 64–66.

[26]. Singh, G.N. (2003): On the improvement of product method of estimation in samplesurveys. Jour. Ind. Soc. Agric. Statisics, 56(3), 267–275.

[27]. Telford, R. D. & Cunningham, R. B. (1991). Sex, sport and body-size dependency of hematology inhighly trained athletes. Med. & Sc. in Sports and Exercise 23: 788-794.