Thailand Statistician
July 2024; 22(3): 634-656
http://statassoc.or.th
Contributed paper

New Product Estimators for Population Mean Under Unequal Probability Sampling with Missing Data: A Case Study on the Number of New COVID-19 Patients

Chugiat Ponkaew [a] and Nuanpan Lawson* [b]
[a] Department of Mathematics and Data Science, Faculty of Science and Technology, Phetchabun Rajabhat University, Phetchabun, Thailand.
[b] Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangkok, Thailand.
*Corresponding author; e-mail: nuanpan.n@sci.kmutnb.ac.th

Received: 8 October 2023
Revised: 24 January 2024
Accepted: 18 February 2024

Abstract

The coronavirus pandemic or COVID-19 has killed numerous human lives and the number of COVID-19 patients all over the world including Thailand has drastically increased. Estimating the incidence of COVID-19 can assist in preventing further impacts through policies and planning for the whole nation. Although some information about COVID-19 are missing. If analysis is conducted without dealing with the issue, imprecise estimations may be made from the data. New product estimators along with the variance estimators for estimating population mean have been introduced under unequal probability sampling without replacement with missing data in the study variable under two nonresponse mechansims; missing completely at random and missing at random. Two frameworks are considered; the two-phase and reverse frameworks to find the variance estimators. Simulation studies and an application to COVID-19 patients investigate the performance of the proposed estimators. The results show that the proposed estimators under the missing at random nonresponse mechanism performs the best with the smallest variance compared to other estimators under both frameworks with the estimated mean of new COVID-19 patients equal to 306 cases per week.

Keywords: Product estimator, COVID-19, nonresponse, reverse framework, unequal probability sampling.

1. Introduction

In sample surveys where there is available information on an auxiliary variable that is related to a study variable, the ratio and product estimators are popular to assist in increasing the efficiency of the population mean estimator. Cochran (1977) invented the ratio estimator for when the auxiliary and study variable are positively related. On the other hand, Murthy (1964) suggested the product estimator when the relationship between these two variables is negative under simple random sampling without replacement (SRSWOR). Under unequal probability sampling without replacement (UPWOR), Horvitz and Thompson's estimator (1952) is renowned for estimating population mean using the first inclusion probability. However, when missing data occur in the study, these estimators cannot be applied to estimate the population mean or population total. Ponkaew and Lawson (2018) suggested a ratio estimator for estimating population mean under UPWOR when the response probability is known or unknown. They considered situations when the sampling fraction is small and can be ignored when the nonresponse mechanism is the missing completely at random (MCAR) under the reverse framework. Ponkaew and Lawson (2019) suggested the population total estimators in the forms of linear and ratio estimators under UPWOR. Their estimators are asymptotically unbiased estimators under the reverse framework when the sampling fraction is omitted. Ponkaew and Lawson (2022) also suggested new ratio estimators by extending the Ponkaew and Lawson (2018) estimator when the population mean of an auxiliary variable is known and unknown and missing data are in both study and auxiliary variables. They investigated their estimators under the reverse framework and the nonresponse mechanisms are MCAR and missing at random (MAR). Ponkaew and Lawson (2024) proposed two new variance estimators by estimating the joint inclusion probability and one that does not require the joint inclusion probability under UPWOR to solve the issue occuring in estimating the variance of the ratio and generalized regression estimators. They considered the uniform nonresponse mechanism in their study and applied the suggested estimators in the simulation studies and an application to fine particulate matter in Thailand. The variance estimator with free joint inclusion probability performs the best giving a narrower confidence interval.

The coronavirus or COVID-19 is a harmful virus that was first discovered in an outbreak in Wuhan, China and has spread throughout the world which causes deaths and severe injury to human respiratory systems. Patients diagnosed with COVID-19 infection will have moderate to severe respiratory problems. Symptoms include fever, cough, and difficulty breathing and may appear within 2-14 days of exposure which is more severe in infants and may also cause very severe symptoms in the elderly. The transmission of the coronavirus is airborne, by coughing and sneezing, or via contact with the contagious agent by close physical contact, touching objects or surfaces that have the virus on them then pass them to your mouth, nose, or eyes before washing your hands. Research is still currently ongoing to understand its mutation capacity, severity and other characteristics of the virus. Although researchers have invented vaccinations to reduce chances of severe infection by the virus and reduce the chance of deaths caused by the pandemic, numerous COVID-19 incidence still increases daily. Estimating the number of new cases could be helpful for the World Health Organization in planning to prepare and prevent severe situations with increasing of COVID-19 patients and mortality. To get precise information to be able to deal with this situation, the lost COVID-19 data that may occur in between the collection procedure must be considered before further analysis.

This paper aims to propose the product estimators when data are missing in the study variable under both MCAR and MAR which allow for non-uniformly nonresponse mechanism under UPWOR The variance estimators of the proposed product estimators are also investigated under the two-phase and reverse frameworks. The performance of the estimators are studied based on simulation studies and an application to new COVID-19 cases in Thailand.

2. Materials and Methods

2.1. Basic setup

The UPWOR is considered in this study. Let y be the study variable, x be the auxiliary variable where y and x are negatively correlated. Let k be a size variable which is correlated with y and is used to define the first and joint inclusion probabilities. Let $U = \{1, 2, ..., N\}$ be the finite population of size N and s be a set of sample size n selected from population U with unequal probability sampling without replacement. Let $\mathscr{F}$ be the set of all possible subsets of U and the sampling design $P(\bullet)$ is the probability measure on the possible s, i.e., $P(s) \geq 0$ for all $s \in \mathscr{F}$. For all i and $j \in U$ we defined the notation as follows, $\pi_i = P(i \in s) = \sum_{s \ni i} P(s)$ is the first order inclusion probability and $\pi_{ij} = P(i \wedge j \in s) = \sum_{s \supset \{i,j\}} P(s)$ is the second order inclusion probability. Let r_i be a response indicator variable of y_i where $r_i = 1$ if unit i responds to item y_i otherwise $r_i = 0$. Let $\mathrm{R} = (r_1 \quad r_2 \quad \cdots \quad r_N)'$ be the vector of the response indicator and $p_i = P(r_i = 1)$ be the response probability under the missing at random (MAR) mechanism while $p = P(r_i = 1)$ is the response probability under the missing completely at random (MCAR) mechanism. Let $E_q(\bullet)$ and $V_q(\bullet)$ be the expectation and variance operators with respect to the nonresponse mechanism. Under a two-phase framework, for all $i \in s$ let t_i be a sample indicator variable of y_i which $t_i = 1$ if unit i is selected in sample s and responds to item y and $p_i = P(t_i = 1)$ and responses to item y. Let $E_p(\bullet)$ and $V_p(\bullet)$ be the expectation and variance operators with respect to sampling design. The overall expectation and variance operators are defined by $E(\bullet)$ and $V(\bullet)$, respectively.

Under the two-phase framework, the expected value and the variance of $\hat{\bar{Y}}$ are equal to

$$E(\hat{\bar{Y}}) = E_p V_q\left(\hat{\bar{Y}} \middle| s\right), \quad V(\hat{\bar{Y}}) = E_p V_q\left(\hat{\bar{Y}} \middle| s\right) + V_p E_q\left(\hat{\bar{Y}} \middle| s\right).$$

Under the reverse framework, the expected value and the variance of $\hat{\bar{Y}}$ can be obtained by,

$$E(\hat{\bar{Y}}) = E_q V_p\left(\hat{\bar{Y}} \middle| \mathrm{R}\right), \quad V(\hat{\bar{Y}}) = E_q V_p\left(\hat{\bar{Y}} \middle| \mathrm{R}\right) + V_q E_p\left(\hat{\bar{Y}} \middle| \mathrm{R}\right).$$

2.2. Existing estimators

2.2.1 The product estimators in the full response

Murthy (1964) introduced a product estimator for estimating population mean when the parameter $\bar{X}$ is known under SRS defined by, rgwi

$$\hat{\bar{Y}}_{PSRS} = \frac{\overline{yx}}{\bar{X}} = \frac{\hat{\bar{P}}_{SRS}}{\bar{X}}, \tag{1}$$

where $\hat{\bar{P}}_{SRS} = \overline{yx}$.

Under UPWOR, Horvitz and Thompson (1952) suggested a population mean estimator for estimating the study variable and it is defined by $\hat{\bar{Y}}_{HT} = \frac{1}{N}\sum_{i \in s}\frac{y_i}{\pi_i}$. The population mean estimator for the auxiliary variable is equal to $\hat{\bar{X}}_{HT} = \frac{1}{N}\sum_{i \in s}\frac{x_i}{\pi_i}$. Then, the modified Murthy (1964) to estimate population mean is

$$\hat{\bar{Y}}_{PHT} = \frac{\hat{\bar{Y}}_{HT}\hat{\bar{X}}_{HT}}{\bar{X}} = \frac{\hat{\bar{P}}_{PHT}}{\bar{X}}, \tag{2}$$

where $\hat{\bar{P}}_{PHT} = \hat{\bar{Y}}_{HT}\hat{\bar{X}}_{HT}$. The properties of $\hat{\bar{Y}}_{PHT}$ are discussed in Lemma 1.

Lemma 1. *Assume that a sample set s of size n is selected according to UPWOR from a population U of size N. Let y and x be the study and auxiliary variables which are negatively correlated and $\hat{\bar{Y}}_{PHT} = \frac{\hat{\bar{Y}}_{HT}\hat{\bar{X}}_{HT}}{\bar{X}} = \frac{\hat{\bar{P}}_{PHT}}{\bar{X}}$ is the product estimator for estimating population mean.*

(1) $\hat{\bar{Y}}_{PHT}$ is an almost unbiased estimator of $\bar{Y}$.

(2) The variance of $\hat{\bar{Y}}_{PHT}$ is

$$V(\hat{\bar{Y}}_{PHT}) = \frac{1}{N^2}\left[\sum_{i\in U} F_i(y_i + Rx_i) + \sum_{i\in U}\sum_{j\setminus\{i\}\in U} J_{ij}(y_i + Rx_i)(y_j + Rx_j)\right],$$

where $F_i = \frac{1-\pi_i}{\pi_i}$, $J_i = \frac{\pi_{ij} - \pi_i\pi_j}{\pi_i\pi_j}$ and $R = \overline{YX}^{-1}$.

(3) The variance estimator of $\hat{\bar{Y}}_{PHT}$ is

$$\hat{V}(\hat{\bar{Y}}_{PHT}) = \frac{1}{N^2}\left[\sum_{i\in s} \hat{F}_i(y_i + \hat{R}_{HT}x_i) + \sum_{i\in s}\sum_{j\setminus\{i\}\in s} \hat{J}_{ij}(y_i + \hat{R}_{HT}x_i)(y_j + \hat{R}_{HT}x_j)\right],$$

where $\hat{F}_i = \frac{1-\pi_i}{\pi_i^2}$, $\hat{J}_i = \frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}\pi_i\pi_j}$ and $\hat{R}_{HT} = \hat{\bar{Y}}_{HT}\hat{\bar{X}}_{HT}^{-1}$.

We noted that if nonresponse occurs in the study variable in other words some units i in sample s cannot be observed then the modified Murthy's (1964) estimator based on the response sample s_r is

$$\hat{\bar{Y}}'_{PHT} = \frac{\hat{\bar{Y}}'_{HT}\hat{\bar{X}}_{HT}}{\bar{X}} = \frac{\hat{\bar{P}}'_{PHT}}{\bar{X}}, \tag{3}$$

where $\hat{\bar{Y}}'_{HT} = \frac{1}{N}\sum_{i\in s_r}\frac{y_i}{\pi_i}$, $\hat{\bar{P}}'_{PHT} = \hat{\bar{Y}}'_{HT}\hat{\bar{X}}_{HT}$. Furthermore, the variance estimator of $\hat{\bar{Y}}'_{PHT}$ is

$$\hat{V}(\hat{\bar{Y}}'_{PHT}) = \frac{1}{N^2}\left[\sum_{i\in s_r} \hat{F}_i(y_i + \hat{R}'_{HT}x_i) + \sum_{i\in s_r}\sum_{j\setminus\{i\}\in s_r} \hat{J}_{ij}(y_i + \hat{R}'_{HT}x_i)(y_j + \hat{R}'_{HT}x_j)\right], \tag{4}$$

where $\hat{F}_i = \frac{1-\pi_i}{\pi_i^2}$, $\hat{J}_i = \frac{\pi_{ij} - \pi_i\pi_j}{\pi_{ij}\pi_i\pi_j}$ and $\hat{R}'_{HT} = \hat{\bar{Y}}'_{HT}\hat{\bar{X}}_{HT}^{-1}$.

2.2.2 The estimators in the presence of nonresponse

In the presence of nonresponse, Ponkaew and Lawson (2019) discussed the population mean estimator under the MCAR mechanism and is defined by

$$\hat{\bar{Y}}'_{PL} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}. \tag{5}$$

If p is unknown this value can be estimated by $\hat{p}=\left(\sum_{i\in s}\frac{r_i}{\pi_i}\right)\left(\sum_{i\in s}\frac{1}{\pi_i}\right)^{-1}$ then the estimator of Ponkaew and Lawson (2019) is given by $\hat{\bar{Y}}'_{PL}=\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i\hat{p}}$. In Lemma 2, we discussed the properties of Ponkaew and Lawson's (2019) estimator under the reverse framework.

Lemma 2. *Assume that a sample set s of size n is selected according to UPWOR from a population U of size N. Let y be the study variable and $\hat{\bar{Y}}'_{PL}=\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}$ is the estimator of $\bar{Y}$. Under the reverse framework with the MCAR mechanism,*

(1) $E(\hat{\bar{Y}}'_{PL})=E_qE_p\left(\hat{\bar{Y}}'_{PL}\middle|\mathrm{R}\right)=\bar{Y}$ then $\hat{\bar{Y}}'_{PL}$ is an unbiased estimator of $\bar{Y}$.

(2) The variance of $\hat{\bar{Y}}'_{PL}$ is

$$V'(\hat{\bar{Y}}'_{PL})\cong\frac{1}{N^2}\left[\sum_{i\in U}F_i y_i^2+\sum_{i\in U}\sum_{j\backslash\{i\}\in U}J_{ij}y_iy_j+\sum_{i\in U}E'_i y_i^2\right],$$

where $F_i=\frac{(1-\pi_i)}{\pi_i}$, $J_{ij}=\frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_i\pi_j}$ and $E'_i=\frac{(1-p)}{p}$.

(3) If p is known then the estimator of $V'(\hat{\bar{Y}}'_{PL})$ is

$$\hat{V}'_p(\hat{\bar{Y}}'_{PL})=\frac{1}{N^2}\left[\sum_{i\in s}\frac{r_i}{p}\hat{F}_i y_i^2+\sum_{i\in s}\sum_{i\backslash\{j\}\in s}\frac{r_i}{p}\frac{r_j}{p}\hat{J}_{ij}y_iy_j+\sum_{i\in s}\frac{r_i}{p}\hat{E}'_i y_i^2\right],$$

where $\hat{F}_i=\frac{(1-\pi_i)}{\pi_i^2}$, $\hat{J}_{ij}=\frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_{ij}\pi_i\pi_j}$ and $\hat{E}'_i=\frac{(1-p)}{p\pi_i}$.

(4) If p is unknown then the estimator of $V'(\hat{\bar{Y}}'_{PL})$ is

$$\hat{V}'_{\hat{p}}(\hat{\bar{Y}}'_{PL})=\frac{1}{N^2}\left[\sum_{i\in s}\frac{r_i}{\hat{p}}\hat{F}_i y_i^2+\sum_{i\in s}\sum_{i\backslash\{j\}\in s}\frac{r_i}{\hat{p}}\frac{r_j}{\hat{p}}\hat{J}_{ij}y_iy_j+\sum_{i\in s}\frac{r_i}{\hat{p}}\hat{\hat{E}}'_i y_i^2\right],$$

where $\hat{F}_i=\frac{(1-\pi_i)}{\pi_i^2}$, $\hat{J}_{ij}=\frac{(\pi_{ij}-\pi_i\pi_j)}{\pi_{ij}\pi_i\pi_j}$, $\hat{\hat{E}}'_i=\frac{(1-\hat{p})}{\hat{p}\pi_i}$ and $\hat{p}=\sum_{i\in s}\frac{r_i}{\pi_i}\left(\sum_{i\in s}\frac{1}{\pi_i}\right)^{-1}$ is the estimator of p.

Then, the properties of $\hat{\bar{Y}}'_{PL}$ under two-phase framework are discussed in Lemma 3.

Lemma 3. *Assume that a sample set s of size n is selected according to UPWOR from a population U of size N. Let y be a study variable and $\hat{\bar{Y}}'_{PL}=\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}$ is the estimator of $\bar{Y}$. Under the two-phase framework with the MCAR mechanism,*

(1) $E\left(\hat{\bar{Y}}'_{PL}\right)=E_pE_q\left(\hat{\bar{Y}}'_{PL}\middle|s\right)=\bar{Y}$ then $\hat{\bar{Y}}'_{PL}$ is an unbiased estimator of $\bar{Y}$.

(2) The variance of $\hat{\bar{Y}}'_{PL}$ is

$$V''(\hat{\bar{Y}}'_{PL}) = \frac{1}{N^2}\left[\sum_{i\in U} F_i y_i^2 + \sum_{i\in U}\sum_{j\setminus\{i\}\in U} J_{ij} y_i y_j + \sum_{i\in U} E'_i \frac{y_i^2}{\pi_i}\right].$$

(3) If p is known then the estimator of $V''(\hat{\bar{Y}}'_{PL})$ is

$$\hat{V}''_p(\hat{\bar{Y}}'_{PL}) = \frac{1}{N^2}\left[\sum_{i\in s} \frac{r_i}{p}\hat{F}_i y_i^2 + \sum_{i\in s}\sum_{j\setminus\{i\}\in s} \frac{r_i}{p}\frac{r_j}{p}\hat{J}_{ij} y_i y_j + \sum_{i\in s} \frac{r_i}{p}\hat{E}'_i \frac{y_i^2}{\pi_i}\right].$$

(4) If p is unknown then the estimator of $V''(\hat{\bar{Y}}'_{PL})$ is

$$\hat{V}''_{\hat{p}}(\hat{\bar{Y}}'_{1PL}) = \frac{1}{N^2}\left[\sum_{i\in s} \frac{r_i}{\hat{p}}\hat{F}_i y_i^2 + \sum_{i\in s}\sum_{j\setminus\{i\}\in s} \frac{r_i}{\hat{p}}\frac{r_j}{\hat{p}}\hat{J}_{ij} y_i y_j + \sum_{i\in s} \frac{r_i}{\hat{p}}\hat{\hat{E}}'_i \frac{y_i^2}{\pi_i}\right],$$

where $\hat{p}$ is the estimator of p.

Ponkaew and Lawson's (2019) estimator in (5) is considered under the MCAR mechanism which is a strong assumption and rarely happens in practice. Ponkaew and Lawson (2019) also discussed a general form of population mean estimator in the presence of nonresponse defined by

$$\hat{\bar{Y}}''_{PL} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p_i}. \tag{6}$$

If p_i is unknown this value can be estimated by $\hat{p}_i$ which obtained by using logistic regression model then the estimator of Ponkaew and Lawson (2019) is given by $\hat{\bar{Y}}''_{PL} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i \hat{p}_i}$. The properties of $\hat{\bar{Y}}''_{PL}$ under the reverse framework with the MAR mechanism are discussed in Lemma 4.

Lemma 4. *Assume that a sample set s of size n is selected according to UPWOR from a population U of size N. Let y be the study variable and $\hat{\bar{Y}}''_{PL} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p_i}$ is the estimator of $\bar{Y}$. Under the reverse framework when the nonresponse mechanism is MAR,*

(1) $E(\hat{\bar{Y}}''_{PL}) = E_q E_p\left(\hat{\bar{Y}}''_{PL}\middle|\mathrm{R}\right) = \bar{Y}$ then $\hat{\bar{Y}}''_{PL}$ is an unbiased estimator of $\bar{Y}$.

(2) The variance of $\hat{\bar{Y}}''_{PL}$ is

$$V'(\hat{\bar{Y}}''_{PL}) \cong \frac{1}{N^2}\left[\sum_{i\in U} F_i y_i^2 + \sum_{i\in U}\sum_{j\setminus\{i\}\in U} J_{ij} y_i y_j + \sum_{i\in U} E''_i y_i^2\right],$$

where $E''_i = \frac{(1-p_i)}{p_i}$.

(3) If p_i is known for all $i\in s$ then the estimator of $V'(\hat{\bar{Y}}''_{PL})$ is

$$\hat{V}'_{p_i}(\hat{\bar{Y}}''_{PL}) = \frac{1}{N^2}\left[\sum_{i\in s} \frac{r_i}{p_i}\hat{F}_i y_i^2 + \sum_{i\in s}\sum_{j\setminus\{i\}\in s} \frac{r_i}{p_i}\frac{r_j}{p_j}\hat{J}_{ij} y_i y_j + \sum_{i\in s} \frac{r_i}{p_i}\hat{E}''_i y_i^2\right],$$

where $\hat{E}''_i = \frac{(1-p_i)}{p_i \pi_i}$.

(4) If p_i is unknown for all $i \in s$ then the estimator of $V'(\hat{\bar{Y}}''_{PL})$ is

$$\hat{V}'_{\hat{p}_i} = \frac{1}{N^2}\left[\sum_{i\in s}\frac{r_i}{\hat{p}_i}\hat{F}_i y_i^2 + \sum_{i\in s}\sum_{i\backslash\{j\}\in s}\frac{r_i}{\hat{p}_i}\frac{r_j}{\hat{p}_j}\hat{J}_{ij}y_i y_j + \sum_{i\in s}\frac{r_i}{\hat{p}_i}\hat{\hat{E}}''_i y_i^2\right],$$

where $\hat{\hat{E}}''_i = \dfrac{(1-\hat{p}_i)}{\hat{p}_i \pi_i}$ and $\hat{p}_i$ is the estimator of p_i.

The properties of $\hat{\bar{Y}}''_{PL}$ under the two phase framework are discussed in Lemma 5.

Lemma 5. *Assume that a sample set s of size n is selected according to UPWOR from a population U of size N. Let y be the study variable and $\hat{\bar{Y}}''_{PL} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p_i}$ is the estimator of $\bar{Y}$. Under the two phase framework with the MAR mechanism,*

(1) $E(\hat{\bar{Y}}''_{PL}) = E_p E_q\left(\hat{\bar{Y}}''_{PL}\middle| s\right) = \bar{Y}$ then $\hat{\bar{Y}}''_{PL}$ is an unbiased estimator of $\bar{Y}$.

(2) The variance of $\hat{\bar{Y}}''_{PL}$ is

$$V''(\hat{\bar{Y}}''_{PL}) \cong \frac{1}{N^2}\left[\sum_{i\in U}F_i y_i^2 + \sum_{i\in U}\sum_{j\backslash\{i\}\in U}J_{ij}y_i y_j + \sum_{i\in U}E''_i\frac{y_i^2}{\pi_i}\right].$$

(3) If p_i is known then the estimator of $V''(\hat{\bar{Y}}''_{PL})$ is

$$\hat{V}''_{p_i}(\hat{\bar{Y}}''_{PL}) = \frac{1}{N^2}\left[\sum_{i\in s}\frac{r_i}{p_i}\hat{F}_i y_i^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\frac{r_i}{p_i}\frac{r_j}{p_j}\hat{J}_{ij}y_i y_j + \sum_{i\in s}\frac{r_i}{p_i}\hat{E}''_i\frac{y_i^2}{\pi_i}\right].$$

(4) If p_i is unknown then the estimator of $V''(\hat{\bar{Y}}'_P)$ is

$$\hat{V}''_{\hat{p}_i}(\hat{\bar{Y}}''_{PL}) = \frac{1}{N^2}\left[\sum_{i\in s}\frac{r_i}{\hat{p}_i}\hat{F}_i y_i^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\frac{r_i}{\hat{p}_i}\frac{r_j}{\hat{p}_j}\hat{J}_{ij}y_i y_j \hat{+} \sum_{i\in s}\frac{r_i}{\hat{p}_i}\hat{\hat{E}}''_i\frac{y_i^2}{\pi_i}\right],$$

where $\hat{p}_i$ is the estimator of p_i by using logistic regression model.

When nonresponse occurs in the study variable but the auxiliary variable is available and positively correlated with the study variable. Ponkaew and Lawson (2018) discussed the ratio estimator to estimate population mean under the MCAR mechanism given by,

$$\hat{\bar{Y}}'_{Ra} = \frac{\hat{\bar{Y}}'_{PL}}{\hat{\bar{X}}_{HT}}\bar{X} = \hat{R}'\bar{X}, \tag{7}$$

where $\hat{\bar{Y}}'_{PL} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}$, $\hat{\bar{X}}_{HT} = \frac{1}{N}\sum_{i\in s}\frac{y_i}{\pi_i}$ and $\hat{R}' = \frac{\hat{\bar{Y}}'_{PL}}{\hat{\bar{X}}_{HT}}$. The properties of $\hat{\bar{Y}}'_{Ra}$ under the reverse framework are discussed in Lemma 6.

Lemma 6. *Assume that a sample set s of size n is selected according to UPWOR from a population U of size N. Let y and x be study and auxiliary variables which are positively correlated. Let*

$\hat{\bar{Y}}'_{Ra} = \dfrac{\hat{\bar{Y}}'_{PL}}{\hat{\bar{X}}_{HT}}\bar{X} = \hat{R}'\bar{X}$ *be the ratio estimator under the MCAR mechanism. Under the reverse framework,*

(1) $E(\hat{\bar{Y}}'_{Ra}) = E_q E_p\left(\hat{\bar{Y}}'_{Ra}\middle|\mathrm{R}\right) \cong \bar{Y}$ *then* $\hat{\bar{Y}}'_{Ra}$ *is an almost unbiased estimator of* $\bar{Y}$.

(2) The variance of $\hat{\bar{Y}}'_{Ra}$ *is*

$$V'(\hat{\bar{Y}}'_{Ra}) \cong \frac{1}{N^2}\left[\sum_{i\in U} F_i \underset{\cdot}{y}_i^2 + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} J_{ij}\underset{\cdot}{y}_i\underset{\cdot}{y}_j + \sum_{i\in U} E'_i \underset{\cdot}{y}_i^2\right],$$

where $F_i = \dfrac{(1-\pi_i)}{\pi_i}$, $J_{ij} = \dfrac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}$, $E'_i = \dfrac{(1-p)}{p}$, $\underset{\cdot}{y}_i = y_i - Rx_i$ *and* $R = \dfrac{\bar{Y}}{\bar{X}}$.

(3) If p *is known then the estimator of* $V'(\hat{\bar{Y}}'_{Ra})$ *is*

$$\hat{V}'_p(\hat{\bar{Y}}'_{Ra}) \cong \frac{1}{N^2}\left[\sum_{i\in s} \hat{F}_i \hat{\underset{\cdot}{y}}_i^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s} \hat{J}_{ij}\hat{\underset{\cdot}{y}}_i\hat{\underset{\cdot}{y}}_j + \sum_{i\in s}\frac{r_i}{p}\hat{E}'_i y_i^2\right],$$

where $\hat{\underset{\cdot}{y}}_i = \dfrac{r_i y_i}{p} - \hat{R}'_p x_i$, $\hat{F}_i = \dfrac{(1-\pi_i)}{\pi_i^2}$, $\hat{J}_{ij} = \dfrac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}\pi_i\pi_j}$, $\hat{E}'_i = \dfrac{(1-p)}{p\pi_i}$ *and* $\hat{R}'_p = \hat{\bar{Y}}'_{PL}\big/\hat{\bar{X}}_{HT}$.

(4) If p *is unknown then the estimator of* $V'(\hat{\bar{Y}}'_P)$ *is*

$$\hat{V}'_{\hat{p}}(\hat{\bar{Y}}'_{Ra}) \cong \frac{1}{N^2}\left[\sum_{i\in s} \hat{F}_i \hat{\hat{\underset{\cdot}{y}}}_i^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s} \hat{J}_{ij}\hat{\hat{\underset{\cdot}{y}}}_i\hat{\hat{\underset{\cdot}{y}}}_j + \sum_{i\in s}\frac{r_i}{\hat{p}}\hat{\hat{E}}'_i y_i^2\right],$$

where $\hat{\hat{\underset{\cdot}{y}}}_i = \dfrac{r_i y_i}{\hat{p}} - \hat{R}'_{\hat{p}} x_i$, $\hat{\hat{E}}'_i = \dfrac{(1-\hat{p})}{\hat{p}\pi_i}$, $\hat{R}'_{\hat{p}} = \hat{\hat{\bar{Y}}}'_{PL}\big/\hat{\bar{X}}_{HT}$, $\hat{\hat{\bar{Y}}}'_{PL} = \dfrac{1}{N}\sum_{i\in s}\dfrac{r_i y_i}{\pi_i \hat{p}}$ *and* $\hat{p} = \sum_{i\in s}\dfrac{r_i}{\pi_i}\left(\sum_{i\in s}\dfrac{1}{\pi_i}\right)^{-1}$ *is the estimator of* p.

The properties of $\hat{\bar{Y}}'_{Ra}$ under the two-phase framework are discussed in Lemma 7.

Lemma 7. *Assume that a sample set* s *of size* n *is selected according to UPWOR from a population* U *of size* N. *Let* y *and* x *be study and auxiliary variables which are positively correlated. Let* $\hat{\bar{Y}}'_{Ra} = \dfrac{\hat{\bar{Y}}'_{PL}}{\hat{\bar{X}}_{HT}}\bar{X} = \hat{R}'\bar{X}$ *be the ratio estimator under the MCAR mechanism. Under the two-phase framework,*

(1) $E(\hat{\bar{Y}}'_{Ra}) = E_p E_q\left(\hat{\bar{Y}}'_{Ra}\middle|s\right) \cong \bar{Y}$ *then* $\hat{\bar{Y}}'_{Ra}$ *is an almost unbiased estimator of* $\bar{Y}$.

(2) The variance of $\hat{\bar{Y}}'_{Ra}$ *is,*

$$V''(\hat{\bar{Y}}'_{Ra}) \cong \frac{1}{N^2}\left[\sum_{i\in U} F_i \underset{\cdot}{y}_i^2 + \sum_{i\in U}\sum_{i\backslash\{j\}\in U} J_{ij}\underset{\cdot}{y}_i\underset{\cdot}{y}_j + \sum_{i\in U}\frac{1}{\pi_i} E'_i \underset{\cdot}{y}_i^2\right],$$

where $F_i = \dfrac{(1-\pi_i)}{\pi_i}$, $J_{ij} = \dfrac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}$, $E'_i = \dfrac{(1-p)}{p}$, $\underset{\cdot}{y}_i = y_i - Rx_i$ *and* $R = \dfrac{\bar{Y}}{\bar{X}}$.

(3) If p is known for all $i \in s$ then the estimator of $V''(\hat{\bar{Y}}'_{Ra})$ is

$$\hat{V}''_p(\hat{\bar{Y}}'_{Ra}) \cong \frac{1}{N^2}\left[\sum_{i\in s}\hat{F}_i\hat{\underset{\cdot}{y}}_i^2 + \sum_{i\in s}\sum_{i\setminus\{j\}\in s}\hat{J}_{ij}\hat{\underset{\cdot}{y}}_i\hat{\underset{\cdot}{y}}_j + \sum_{i\in s}\frac{r_i}{p}\frac{1}{\pi_i^2}\hat{E}'_i y_i^2\right],$$

where $\hat{\underset{\cdot}{y}}_i = \dfrac{r_i y_i}{p} - \hat{R}' x_i$, $\hat{F}_i = \dfrac{(1-\pi_i)}{\pi_i^2}$, $\hat{J}_{ij} = \dfrac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}\pi_i\pi_j}$, $\hat{E}'_i = \dfrac{(1-p)}{p\pi_i}$ *and* $\hat{R}' = \hat{\bar{Y}}'_{PL}\big/\hat{\bar{X}}_{HT}$.

(4) If p is unknown for all $i \in s$ then the estimator of $V''(\hat{\bar{Y}}'_P)$ is

$$\hat{V}''_{\hat{p}}(\hat{\bar{Y}}'_{Ra}) \cong \frac{1}{N^2}\left[\sum_{i\in s}\hat{F}_i\hat{\hat{\underset{\cdot}{y}}}_i^2 + \sum_{i\in s}\sum_{i\setminus\{j\}\in s}\hat{J}_{ij}\hat{\hat{\underset{\cdot}{y}}}_i\hat{\hat{\underset{\cdot}{y}}}_j + \sum_{i\in s}\frac{r_i}{\hat{p}}\frac{1}{\pi_i^2}\hat{\hat{E}}'_i y_i^2\right],$$

where $\hat{\hat{\underset{\cdot}{y}}}_i = \dfrac{r_i y_i}{\hat{p}} - \hat{\hat{R}}' x_i$, $\hat{\hat{E}}'_i = \dfrac{(1-\hat{p})}{\hat{p}\pi_i}$, $\hat{\hat{R}}' = \hat{\bar{Y}}'_{PL}\big/\hat{\bar{X}}_{HT}$, $\hat{\bar{Y}}'_{PL} = \dfrac{1}{N}\sum_{i\in s}\dfrac{r_i y_i}{\pi_i \hat{p}}$ *and* $\hat{p} = \sum_{i\in s}\dfrac{r_i}{\pi_i}\left(\sum_{i\in s}\dfrac{1}{\pi_i}\right)^{-1}$ *is the estimator of p.*

Ponkaew and Lawson (2022) suggested the ratio estimator in the presence of nonresponse in study variable under MAR defined by

$$\hat{\bar{Y}}''_{Ra} = \frac{\hat{\bar{Y}}''_{PL}}{\hat{\bar{X}}_{HT}}\bar{X} = \hat{R}''\bar{X}, \tag{8}$$

where $\hat{\bar{Y}}''_{PL} = \dfrac{1}{N}\sum_{i\in s}\dfrac{r_i y_i}{\pi_i p_i}$, $\hat{\bar{X}}_{HT} = \dfrac{1}{N}\sum_{i\in s}\dfrac{y_i}{\pi_i}$ and $\hat{R}'' = \dfrac{\hat{\bar{Y}}''_{PL}}{\hat{\bar{X}}_{HT}}$. The properties of $\hat{\bar{Y}}''_{Ra}$ under the reverse framework are discussed in Lemma 8.

Lemma 8. *Assume that a sample set s of size n is selected according to UPWOR from a population U of size N. Let y and x be study and auxiliary variables which are positively correlated. Let $\hat{\bar{Y}}''_{Ra} = \dfrac{\hat{\bar{Y}}''_{PL}}{\hat{\bar{X}}_{HT}}\bar{X} = \hat{R}''\bar{X}$ be the ratio estimator under the MAR mechanism. Under the reverse framework,*

(1) $E(\hat{\bar{Y}}''_{Ra}) = E_q E_p\left(\hat{\bar{Y}}''_{Ra}\middle|\mathrm{R}\right) \cong \bar{Y}$ then $\hat{\bar{Y}}''_{Ra}$ is an almost unbiased estimator of $\bar{Y}$.

(2) The variance of $\hat{\bar{Y}}''_{Ra}$ is

$$V'(\hat{\bar{Y}}''_{Ra}) \cong \frac{1}{N^2}\left[\sum_{i\in U}F_i\underset{\cdot\cdot}{y}_i^2 + \sum_{i\in U}\sum_{j\setminus\{i\}\in U}J_{ij}\underset{\cdot\cdot}{y}_i\underset{\cdot\cdot}{y}_j + \sum_{i\in U}E''_i y_i^2\right],$$

where $F_i = \dfrac{(1-\pi_i)}{\pi_i}$, $J_{ij} = \dfrac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}$, $E''_i = \dfrac{(1-p_i)}{p_i}$, $\underset{\cdot\cdot}{y}_i = \underset{\cdot}{y}_i = y_i - Rx_i$ *and* $R = \dfrac{\bar{Y}}{\bar{X}}$.

(3) If p_i is known then the estimator of $V'(\hat{\bar{Y}}''_{Ra})$ is

$$\hat{V}'_{p_i}(\hat{\bar{Y}}''_{Ra}) \cong \frac{1}{N^2}\left[\sum_{i\in s}\hat{F}_i\hat{\underset{\cdot\cdot}{y}}_i^2 + \sum_{i\in s}\sum_{i\setminus\{j\}\in s}\hat{J}_{ij}\hat{\underset{\cdot\cdot}{y}}_i\hat{\underset{\cdot\cdot}{y}}_j + \sum_{i\in s}\frac{r_i}{p_i}\hat{E}''_i y_i^2\right],$$

where $\hat{\underset{\cdot\cdot}{y}}_i = \dfrac{r_i y_i}{p_i} - \hat{R}'' x_i$, $\hat{F}_i = \dfrac{(1-\pi_i)}{\pi_i^2}$, $\hat{J}_{ij} = \dfrac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}\pi_i\pi_j}$, $\hat{E}''_i = \dfrac{(1-p_i)}{p_i\pi_i}$ *and* $\hat{R}'' = \hat{\bar{Y}}''_{PL}\big/\hat{\bar{X}}_{HT}$.

(4) If p_i *is unknown then the estimator of* $V'(\hat{\hat{\bar{Y}}}''_{Ra})$ *is*

$$\hat{V}'_{\hat{p}_i}(\hat{\hat{\bar{Y}}}''_{Ra}) \cong \frac{1}{N^2}\left[\sum_{i\in s}\hat{F}_i \hat{\underset{..}{y}}_i^2 + \sum_{i\in s}\sum_{i\backslash\{j\}\in s}\hat{J}_{ij}\hat{\underset{..}{y}}_i \hat{\underset{..}{y}}_j + \sum_{i\in s}\frac{r_i}{\hat{p}_i}\hat{\hat{E}}''_i y_i^2\right],$$

where $\hat{\hat{\underset{..}{y}}}_i = \dfrac{r_i y_i}{\hat{p}_i} - \hat{\hat{R}}'' x_i$, $\hat{E}'_{2i} = \dfrac{(1-\hat{p}_i)}{\hat{p}_i \pi_i}$, $\hat{\hat{R}}'' = \hat{\bar{Y}}''_{PL} \big/ \hat{\bar{X}}_{HT}$, $\hat{\bar{Y}}''_{PL} = \dfrac{1}{N}\sum_{i\in s}\dfrac{r_i y_i}{\pi_i \hat{p}_i}$ *and* $\hat{p}_i$ *is the estimator of* p_i.

The properties of $\hat{\hat{\bar{Y}}}''_{Ra}$ under two-phase framework are discussed in Lemma 9.

Lemma 9. *Assume that a sample set* s *of size* n *is selected according to UPWOR from a population* U *of size* N. *Let* y *and* x *be study and auxiliary variables which are positively related. Let* $\hat{\hat{\bar{Y}}}''_{Ra} = \dfrac{\hat{\hat{\bar{Y}}}''_{PL}}{\hat{\bar{X}}_{HT}}\bar{X} = \hat{R}''\bar{X}$ *be the ratio estimator under the MAR mechanism. Under the two-phase framework,*

(1) $E(\hat{\hat{\bar{Y}}}''_{Ra}) = E_p E_q\left(\hat{\hat{\bar{Y}}}''_{Ra}\middle| \mathrm{s}\right) \cong \bar{Y}$ *then* $\hat{\hat{\bar{Y}}}''_{Ra}$ *is an almost unbiased estimator of* $\bar{Y}$.

(2) The variance of $\hat{\hat{\bar{Y}}}''_{Ra}$ *is*

$$V''(\hat{\hat{\bar{Y}}}''_{Ra}) \cong \frac{1}{N^2}\left[\sum_{i\in U} F_i \underset{..}{y}_i^2 + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} J_{ij}\underset{..}{y}_i \underset{..}{y}_j + \sum_{i\in U}\frac{1}{\pi_i}E''_i y_i^2\right],$$

where $F_i = \dfrac{(1-\pi_i)}{\pi_i}$, $J_{ij} = \dfrac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}$, $E''_i = \dfrac{(1-p_i)}{p_i}$, $\underset{..}{y}_i = y_i - Rx_i$ *and* $R = \dfrac{\bar{Y}}{\bar{X}}$.

(3) If p_i *is known then the estimator of* $V''(\hat{\hat{\bar{Y}}}'_{Ra})$ *is*

$$\hat{V}''_{p_i}(\hat{\hat{\bar{Y}}}''_{Ra}) \cong \frac{1}{N^2}\left[\sum_{i\in s}\hat{F}_i \hat{\underset{..}{y}}_i^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{J}_{ij}\hat{\underset{..}{y}}_i \hat{\underset{..}{y}}_j + \sum_{i\in s}\frac{r_i}{p}\frac{1}{\pi_i^2}\hat{E}''_i y_i^2\right],$$

where $\hat{\underset{..}{y}}_i = \dfrac{r_i y_i}{p_i} - \hat{R}'' x_i$, $\hat{F}_i = \dfrac{(1-\pi_i)}{\pi_i^2}$, $\hat{J}_{ij} = \dfrac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}\pi_i \pi_j}$, $\hat{E}''_i = \dfrac{(1-p_i)}{p_i \pi_i}$ *and* $\hat{R}'' = \hat{\hat{\bar{Y}}}''_{PL} \big/ \hat{\bar{X}}_{HT}$.

(4) If p_i *is unknown then the estimator of* $V''(\hat{\hat{\bar{Y}}}''_{Ra})$ *is*

$$\hat{V}''_{\hat{p}_i}(\hat{\hat{\bar{Y}}}''_{Ra}) \cong \frac{1}{N^2}\left[\sum_{i\in s}\hat{F}_i \hat{\hat{\underset{..}{y}}}_i^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{J}_{ij}\hat{\hat{\underset{..}{y}}}_i \hat{\hat{\underset{..}{y}}}_j + \sum_{i\in s}\frac{r_i}{\hat{p}}\frac{1}{\pi_i^2}\hat{E}''_i y_i^2\right],$$

where $\hat{\hat{\underset{..}{y}}}_i = \dfrac{r_i y_i}{\hat{p}_i} - \hat{\hat{R}}'' x_i$, $\hat{E}''_i = \dfrac{(1-\hat{p})}{\hat{p}\pi_i}$, $\hat{\hat{R}}'' = \hat{\bar{Y}}''_{PL} \big/ \hat{\bar{X}}_{HT}$, $\hat{\bar{Y}}''_{PL} = \dfrac{1}{N}\sum_{i\in s}\dfrac{r_i y_i}{\pi_i \hat{p}_i}$ *and* $\hat{p}_i$ *is the estimator of* p_i.

2.3. The proposed estimators and associated variance estimators

2.3.1 The proposed estimators

Two new product estimators for estimating population mean are proposed when nonresponse occurs in study variable and the population mean $\bar{X}$ is known under MCAR and MAR. Under MCAR, we proposed to develop the product estimator in a full response case and the linear estimator proposed by Ponkaew and Lawson (2019). Recall from (2), the product estimator in the full response case is $\hat{\bar{Y}}_{PHT}=\frac{\hat{\bar{Y}}_{HT}\hat{\bar{X}}_{HT}}{\bar{X}}=\frac{\hat{\bar{P}}_{PHT}}{\bar{X}}$. If the nonresponse mechanism is MCAR the linear estimator of Ponkaew and Lawson (2019) is $\hat{\bar{Y}}'_{PL}=\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}$. We suggested to replace $\hat{\bar{Y}}_{HT}$ in (2) with $\hat{\bar{Y}}'_{PL}$. The proposed product estimator under MCAR is

$$\hat{\bar{Y}}'_P=\frac{\hat{\bar{Y}}'_{PL}\hat{\bar{X}}_{HT}}{\bar{X}}=\frac{\hat{P}'}{\bar{X}}, \tag{9}$$

where $\hat{\bar{Y}}'_{PL}=\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}$, $\hat{\bar{X}}_{HT}=\frac{1}{N}\sum_{i\in s}\frac{y_i}{\pi_i}$ and $\hat{P}'=\hat{\bar{Y}}'_{PL}\hat{\bar{X}}_{HT}$. However, if p is unknow we can estimated by $\hat{p}=\sum_{i\in s}\frac{r_i}{\pi_i}\Big/\sum_{i\in s}\frac{1}{\pi_i}$. In Theorem 1, we showed that the proposed estimator is an almost unbiased estimator of $\bar{Y}$.

Theorem 1. *Assume that a sample set s of size n is selected according to UPWOR from a population U of size N. Let y and x be study and auxiliary variables which are negatively correlated. Let $\hat{\bar{Y}}'_P=\frac{\hat{\bar{Y}}'_{PL}\hat{\bar{X}}_{HT}}{\bar{X}}=\frac{\hat{P}'}{\bar{X}}$ be the new product estimator under the MCAR mechanism*

(1) Under the reverse framework $E(\hat{\bar{Y}}'_P)=E_qE_p\left(\hat{\bar{Y}}'_P\middle|\mathrm{R}\right)\cong\bar{Y}$ then $\hat{\bar{Y}}'_P$ is an almost unbiased estimator of $\bar{Y}$.

(2) Under the two-phase framework $E(\hat{\bar{Y}}'_P)=E_pE_q\left(\hat{\bar{Y}}'_P\middle|s\right)\cong\bar{Y}$ then $\hat{\bar{Y}}'_P$ is an almost unbiased estimator of $\bar{Y}$.

Proof:

(1) We showed that $E(\hat{\bar{Y}}'_P)=E_qE_p\left(\hat{\bar{Y}}'_P\middle|\mathrm{R}\right)\cong\bar{Y}$. Let $\hat{\bar{Y}}'_P=\frac{\hat{\bar{Y}}'_{PL}\hat{\bar{X}}_{HT}}{\bar{X}}=\frac{\hat{P}'}{\bar{X}}$. Under the reverse framework the overall expectation of $\hat{\bar{Y}}'_P$ is defined by,

$$E(\hat{\bar{Y}}'_P)=E_qE_p\left(\hat{\bar{Y}}'_P\middle|\mathrm{R}\right). \tag{10}$$

However, $\hat{\bar{Y}}'_P$ is in the form of a nonlinear function. Then, we use the Taylor linearization approach to transform this estimator into a linear function defined by

$$\hat{\bar{Y}}'_{P,lin}\cong\tilde{\bar{Y}}'_{PL}+(\hat{T}_1-T_1)+\tilde{R}'(\hat{T}_2-T_2), \tag{11}$$

where $\hat{T}_1=\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}$, $T_1=E_p\left(\hat{T}_1\middle|\mathrm{R}\right)=\tilde{\bar{Y}}'_{PL}=\frac{1}{N}\sum_{i\in U}\frac{r_i y_i}{p}$, $\hat{T}_2=\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}$, $T_2=E_p\left(\hat{T}_2\middle|\mathrm{R}\right)=\bar{X}$ and $\tilde{R}'=\frac{\tilde{\bar{Y}}'_{PL}}{\bar{X}}$.

Then, $E(\hat{\bar{Y}}'_P)$ can be approximated from,

$$E(\hat{\bar{Y}}'_P) \cong E_q E_p\left(\hat{\bar{Y}}'_{P,lin}\middle|\mathrm{R}\right) = E_q E_p\left(\tilde{\bar{Y}}'_{PL} + (\hat{T}_1 - T_1) + \tilde{R}'(\hat{T}_2 - T_2)\middle|\mathrm{R}\right) = \bar{Y}.$$

Therefore, $E(\hat{\bar{Y}}'_P) \cong \bar{Y}$ in other words $\hat{\bar{Y}}'_P$ is an almost unbiased estimator of $\bar{Y}$. The proof of item (2) is similar to item (1).

Under MAR, we proposed a new product estimator. Recall from (2), the product estimator in full response is $\hat{\bar{Y}}_{PHT} = \dfrac{\hat{\bar{Y}}_{HT}\hat{\bar{X}}_{HT}}{\bar{X}} = \dfrac{\hat{\bar{P}}_{PHT}}{\bar{X}}$. In the presence of nonresponse and the nonresponse mechanism is MAR the linear estimator of Ponkaew and Lawson (2019) is $\hat{\bar{Y}}''_p = \dfrac{1}{N}\sum_{i\in s}\dfrac{r_i y_i}{\pi_i p_i}$. We suggested to replace $\hat{\bar{Y}}_{HT}$ in (2) with $\hat{\bar{Y}}''_p$. The new product estimator under MAR is

$$\hat{\bar{Y}}''_p = \frac{\hat{\bar{Y}}''_p\hat{\bar{X}}_{HT}}{\bar{X}} = \frac{\hat{P}''}{\bar{X}}, \tag{12}$$

where $\hat{\bar{Y}}''_p = \dfrac{1}{N}\sum_{i\in s}\dfrac{r_i y_i}{\pi_i p_i}$, $\hat{\bar{X}}_{HT} = \dfrac{1}{N}\sum_{i\in s}\dfrac{y_i}{\pi_i}$ and $\hat{P}'' = \hat{\bar{Y}}''_p\hat{\bar{X}}_{HT}$. In (12) if the probability of response p_i is unknown then it can be estimated by $\hat{p}_i$ which obtained by using logistic regression model. The new product estimator under MAR is in the form of a nonlinear function then the expectation of this estimator can be obtained by using the Taylor linearization approach and it is discussed in Theorem 2.

Theorem 2. *Assume that a sample set s of size n is selected according to UPWOR from a population U of size N. Let y and x be study and auxiliary variables which are negatively correlated. Let $\hat{\bar{Y}}''_p = \dfrac{\hat{\bar{Y}}''_p\hat{\bar{X}}_{HT}}{\bar{X}} = \dfrac{\hat{P}''}{\bar{X}}$ be the new product estimator under the MAR mechanism.*

(1) Under the reverse framework $E(\hat{\bar{Y}}''_p) = E_q E_p\left(\hat{\bar{Y}}''_p\middle|\mathrm{R}\right) \cong \bar{Y}$ then $\hat{\bar{Y}}''_p$ is an almost unbiased estimator of $\bar{Y}$.

(2) Under the two-phase framework $E(\hat{\bar{Y}}''_p) = E_p E_q\left(\hat{\bar{Y}}''_p\middle|s\right) \cong \bar{Y}$ then $\hat{\bar{Y}}''_p$ is an almost unbiased estimator of $\bar{Y}$.

Proof: The proof in Theorem 2 is similar to Theorem 1.

2.3.2 The variance and associated estimators of the proposed estimators

We investigate the variance and the associated estimators of two new product estimators under MCAR and MAR both under the reverse and two-phase frameworks. In Theorems 3 and 4, the variance and associated estimators of new product estimators under the MCAR mechanism are discussed.

Theorem 3. *Assume that a sample s of size n is selected according to UPWOR from a population U of size N. Let y and x be study and auxiliary variables which are negatively correlated. Let $\hat{\hat{Y}}'_P = \dfrac{\hat{\hat{Y}}_{PL}\hat{\hat{X}}_{HT}}{\bar{X}} = \dfrac{\hat{P}'}{\bar{X}}$ be the new product estimator when the nonresponse mechanism is MCAR under the reverse framework.*

(1) The variance of $\hat{\hat{Y}}'_P$ is

$$V'(\hat{\hat{Y}}'_P) \cong \frac{1}{N^2}\left[\sum_{i\in U} F_i \dot{y}_i'^2 + \sum_{i\in U}\sum_{j\backslash\{i\}\in U} J_{ij}\dot{y}'_i\dot{y}'_j + \sum_{i\in U} E'_i y_i^2\right],$$

where $F_i = \dfrac{(1-\pi_i)}{\pi_i}$, $J_{ij} = \dfrac{\pi_{ij}-\pi_i\pi_j}{\pi_i\pi_j}$, $E'_i = \dfrac{(1-p)}{p}$, $\dot{y}'_i = y_i + Rx_i$, $\dot{y}'_j = y_j + Rx_j$ *and* $R = \dfrac{\bar{Y}}{\bar{X}}$.

(2) If p is known then the estimator of $V'(\hat{\hat{Y}}'_P)$ is

$$\hat{V}'_p(\hat{\hat{Y}}'_P) \cong \frac{1}{N^2}\left[\sum_{i\in s} \hat{F}_i \hat{\dot{y}}_i'^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s} \hat{J}_{ij}\hat{\dot{y}}'_i\hat{\dot{y}}'_j + \sum_{i\in s} \frac{r_i}{p}\hat{E}'_i y_i^2\right],$$

where $\hat{\dot{y}}'_i = \dfrac{r_i y_i}{p} + \hat{R}'x_i$, $\hat{\dot{y}}'_j = \dfrac{r_j y_j}{p} + \hat{R}'x_j$, $\hat{F}_i = \dfrac{(1-\pi_i)}{\pi_i^2}$, $\hat{J}_{ij} = \dfrac{\pi_{ij}-\pi_i\pi_j}{\pi_{ij}\pi_i\pi_j}$, $\hat{E}'_{1i} = \dfrac{(1-p)}{p\pi_i}$,

$\hat{\hat{Y}}'_{PL} = \dfrac{1}{N}\sum_{i\in s}\dfrac{r_i y_i}{\pi_i p}$, $\hat{\hat{X}}_{HT} = \dfrac{1}{N}\sum_{i\in s}\dfrac{x_i}{\pi_i}$ *and* $\hat{R}' = \hat{\hat{Y}}'_{PL}\big/\hat{\hat{X}}_{HT}$.

(3) If p is unknown then the estimator of $V'(\hat{\hat{Y}}'_P)$ is

$$\hat{V}'_{\hat{p}} \cong \frac{1}{N^2}\left[\sum_{i\in s} \hat{F}_i \hat{\hat{\dot{y}}}_i'^2 + \sum_{i\in s}\sum_{j\backslash\{i\}\in s} \hat{J}_{ij}\hat{\hat{\dot{y}}}'_i\hat{\hat{\dot{y}}}'_j + \sum_{i\in s} \frac{r_i}{\hat{p}}\hat{\hat{E}}'_i y_i^2\right],$$

where $\hat{\hat{Y}}'_{PL} = \dfrac{1}{N}\sum_{i\in s}\dfrac{r_i y_i}{\pi_i \hat{p}}$, $\hat{\hat{X}}_{HT} = \dfrac{1}{N}\sum_{i\in s}\dfrac{x_i}{\pi_i}$, $\hat{\hat{R}}' = \hat{\hat{Y}}'_{PL}\big/\hat{\hat{X}}_{HT}$, $\hat{\hat{\dot{y}}}'_i = \dfrac{r_i y_i}{\hat{p}} + \hat{R}'x_i$, $\hat{\hat{\dot{y}}}'_j = \dfrac{r_j y_j}{\hat{p}} + \hat{R}'x_j$,

$\hat{\hat{E}}'_i = \dfrac{(1-\hat{p})}{\hat{p}\pi_i}$ *and* $\hat{p} = \sum_{i\in s}\dfrac{r_i}{\pi_i}\left(\sum_{i\in s}\dfrac{1}{\pi_i}\right)^{-1}$ *is the estimator of p.*

Proof: Let $\hat{\hat{Y}}'_P = \dfrac{\hat{\hat{Y}}'_{PL}\hat{\hat{X}}_{HT}}{\bar{X}} = \dfrac{\hat{P}'}{\bar{X}}$.

(1) Under reverse framework the variance of $\hat{\hat{Y}}'_P$ can be obtained by

$$V'(\hat{\hat{Y}}'_P) = E_q V_p\left(\hat{\hat{Y}}'_P\middle|\mathrm{R}\right) + V_q E_p\left(\hat{\hat{Y}}'_P\middle|\mathrm{R}\right). \tag{13}$$

We investigated the term $E_q V_p\left(\hat{\hat{Y}}'_P\middle|\mathrm{R}\right)$ in (13). Since $\hat{\hat{Y}}'_P$ is in a nonlinear form then we use the Taylor linearization approach to transform this value to a linear function given by

$$\hat{\hat{Y}}'_{P,lin} \cong \tilde{\bar{Y}}'_{PL} + (\hat{T}_1 - T_1) + \tilde{R}'(\hat{T}_2 - T_2), \tag{14}$$

where $\hat{T}_1 = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}$, $T_1 = E_p\left(\hat{T}_1\middle|\mathrm{R}\right) = \tilde{\bar{Y}}'_{PL} = \frac{1}{N}\sum_{i\in U}\frac{r_i y_i}{p}$, $\hat{T}_2 = \frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}$, $T_2 = E_p\left(\hat{T}_2\middle|\mathrm{R}\right) = \bar{X}$ and $\tilde{R}' = \frac{\tilde{\bar{Y}}'_{PL}}{\bar{X}}$. From (14), we may rewrite $\hat{\bar{Y}}'_{P,lin}$ as $\hat{\bar{Y}}'_{P,lin} \cong C + \frac{1}{N}\sum_{i\in s}\frac{\tilde{\dot{y}}'_i}{\pi_i}$ where C is a constant $\tilde{\dot{y}}'_i = \frac{r_i y_i}{p} + \tilde{R}' x_i$. Then, $E_q V_p\left(\hat{\bar{Y}}'_P\middle|\mathrm{R}\right)$ can be approximated by

$$
\begin{aligned}
E_q V_p\left(\hat{\bar{Y}}'_P\middle|\mathrm{R}\right) &\cong E_q V_p\left(\hat{\bar{Y}}'_{P,lin}\middle|\mathrm{R}\right) = E_q V_p\left(C + \frac{1}{N}\sum_{i\in U}\frac{\tilde{\dot{y}}'_i}{\pi_i}\middle|\mathrm{R}\right) \\
&= \frac{1}{N^2}E_q\left(\sum_{i\in U}F_i\tilde{\dot{y}}'^2_i + \sum_{i\in U}\sum_{j\backslash\{i\}\in U}J_{ij}\tilde{\dot{y}}'_i\tilde{\dot{y}}'_j\middle|\mathrm{R}\right) \\
&\cong \frac{1}{N^2}\left[\sum_{i\in U}F_i\dot{y}'^2_i + \sum_{i\in U}\sum_{j\backslash\{i\}\in U}J_{ij}\dot{y}'_i\dot{y}'_j\right],
\end{aligned}
$$

where $E_q\left(\tilde{\dot{y}}'_i\middle|\mathrm{R}\right) \cong \dot{y}'_i = y_i + Rx_i$. Therefore,

$$
E_q V_p\left(\hat{\bar{Y}}'_P\middle|\mathrm{R}\right) \cong \frac{1}{N^2}\left[\sum_{i\in U}F_i\dot{y}'^2_i + \sum_{i\in U}\sum_{j\backslash\{i\}\in U}J_{ij}\dot{y}'_i\dot{y}'_j\right]. \tag{15}
$$

Then, we investigated the term $V_q E_p\left(\hat{\bar{Y}}'_P\middle|\mathrm{R}\right)$,

$$
\begin{aligned}
V_q E_p\left(\hat{\bar{Y}}'_P\middle|\mathrm{R}\right) &= V_q E_p\left(\frac{\hat{\bar{Y}}'_{PL}\hat{\bar{X}}_{HT}}{\bar{X}}\middle|\mathrm{R}\right) = V_q E_p\left(\frac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}{\bar{X}}\middle|\mathrm{R}\right) \\
&\cong V_q\left(\frac{\frac{1}{N}\sum_{i\in U}\frac{r_i y_i}{p}\bar{X}}{\bar{X}}\middle|\mathrm{R}\right) = V_q\left(\frac{1}{N}\sum_{i\in U}\frac{r_i y_i}{p}\middle|\mathrm{R}\right) = \frac{1}{N^2}\sum_{i\in U}\frac{(1-p)}{p}y_i^2 \\
&= \frac{1}{N^2}\sum_{i\in U}E'_i y_i^2,
\end{aligned}
$$

where $E'_i = \frac{1-p}{p}$. Therefore,

$$
V_q E_p\left(\hat{\bar{Y}}'_P\middle|\mathrm{R}\right) \cong \frac{1}{N^2}\sum_{i\in U}E'_i y_i^2. \tag{16}
$$

Substitute (15) and (16) in (13) we have,

$$
V'(\hat{\bar{Y}}'_P) \cong \frac{1}{N^2}\left[\sum_{i\in U}F_i\dot{y}'^2_i + \sum_{i\in U}\sum_{j\backslash\{i\}\in U}J_{ij}\dot{y}'_i\dot{y}'_j + \sum_{i\in U}E'_i y_i^2\right]. \tag{17}
$$

(2) Assume that p is known. The estimator of $V'(\hat{\bar{Y}}'_P)$ can be obtained by estimating the unknown parameters in (17) and it is equal to

$$
\hat{V}'_p(\hat{\bar{Y}}'_P) \cong \frac{1}{N^2}\left[\sum_{i\in s}\hat{F}_i\hat{\dot{y}}'^2_i + \sum_{i\in s}\sum_{j\backslash\{i\}\in s}\hat{J}_{ij}\hat{\dot{y}}'_i\hat{\dot{y}}'_j + \sum_{i\in s}\frac{r_i}{p}\hat{E}'_i y_i^2\right].
$$

(3) Assume that p is unknown. Under the MCAR mechanism it can be estimated by $\hat{p}=\sum_{i\in s}\frac{r_i}{\pi_i}\left(\sum_{i\in s}\frac{1}{\pi_i}\right)^{-1}$. Therefore, the variance estimator of $\hat{\bar{Y}}_P'$ is

$$\hat{V}_{\hat{p}}'(\hat{\bar{Y}}_P')\cong\frac{1}{N^2}\left[\sum_{i\in s}\hat{F}_i\hat{\dot{y}}_{1i}'^2+\sum_{i\in s}\sum_{j\setminus\{i\}\in s}\hat{J}_{ij}\hat{\dot{y}}_{1i}'\hat{\dot{y}}_{1j}'+\sum_{i\in s}\frac{r_i}{\hat{p}}\hat{E}_i'y_i^2\right].$$

Next in Theorem 4, we discussed the variance of the new product estimator along with their estimators under the two-phase framework.

Theorem 4. *Assume that a sample set s of size n is selected according to UPWOR from a population U of size N. Let y and x be study and auxiliary variables which are negatively related. Let $\hat{\bar{Y}}_P'=\frac{\hat{\bar{Y}}_{PL}'\hat{\bar{X}}_{HT}}{\bar{X}}=\frac{\hat{P}'}{\bar{X}}$ be the new product estimator when the nonresponse mechanism is MCAR. Under two-phase framework.*

(1) The variance of $\hat{\bar{Y}}_P'$ is

$$V''(\hat{\bar{Y}}_P')\cong\frac{1}{N^2}\left[\sum_{i\in U}F_i\dot{y}_i'^2+\sum_{i\in U}\sum_{j\setminus\{i\}\in U}J_{ij}\dot{y}_i'\dot{y}_j'+\sum_{i\in U}\frac{1}{\pi_i}E_i'y_i^2\right].$$

(2) If p is known then the estimator of $V''(\hat{\bar{Y}}_P')$ is

$$\hat{V}_p''(\hat{\bar{Y}}_P')\cong\frac{1}{N^2}\left[\sum_{i\in s}\hat{F}_i\hat{\dot{y}}_i'^2+\sum_{i\in s}\sum_{j\setminus\{i\}\in s}\hat{J}_{ij}\hat{\dot{y}}_i'\hat{\dot{y}}_j'+\sum_{i\in s}\frac{r_i}{p}\frac{1}{\pi_i}\hat{E}_i'y_i^2\right].$$

(3) If p is unknown then the estimator of $V''(\hat{\bar{Y}}_P')$ is

$$\hat{V}_{\hat{p}}''(\hat{\bar{Y}}_P')\cong\frac{1}{N^2}\left[\sum_{i\in s}\hat{F}_i\hat{\dot{y}}_{1i}'^2+\sum_{i\in s}\sum_{j\setminus\{i\}\in s}\hat{J}_{ij}\hat{\dot{y}}_{1i}'\hat{\dot{y}}_{1j}'+\sum_{i\in s}\frac{r_i}{\hat{p}}\frac{1}{\pi_i}\hat{E}_i'y_i^2\right],$$

where $\hat{p}=\sum_{i\in s}\frac{r_i}{\pi_i}\left(\sum_{i\in s}\frac{1}{\pi_i}\right)^{-1}$.

Proof: The proof of Theorem 4 is similar to Theorem 3.

From Theorem 3 and Theorem 4, we see that the variance of the new product estimators under the reverse framework and the two-phase framework are very similar except two terms; $\sum_{i\in U}\frac{1}{\pi_i^2}E_i'y_i^2$ and $\sum_{i\in U}E_i'y_i^2$. Next, in Lemma 10 we showed that the variance of the new product estimators when considering it under the reverse framework is always smaller than the variance under the two-phase framework.

Lemma 10. *Under the MCAR mechanism. If $p\neq 0$ and $\pi_i\neq 0$ for all $i\in U$ then $V'(\hat{\bar{Y}}_P')\leq V''(\hat{\bar{Y}}_P')$.*

Proof: Let $p \neq 0$ and $\pi_i \neq 0$ for all $i \in U$. Since $0 \leq p \leq 1$ but $p \neq 0$ then $0 < p \leq 1$, $0 \leq 1-p < 1$ and $p^{-1} > 1$. Therefore, we can conclude that $E_i' = (1-p)p^{-1} \geq 0$. Similarly, for π_i, we can conclude that $\pi_i^{-1} \geq 1$. The value of $E_i' \leq \frac{1}{\pi_i} E_i'$ for all $i \in U$. We showed that $V'(\hat{\bar{Y}}_P') \leq V''(\hat{\bar{Y}}_P')$. We have $E_i' \leq \frac{1}{\pi_i} E_i'$ for all $i \in U$ then,

$$\sum_{i \in U} E_i y_i^2 \leq \sum_{i \in U} \frac{1}{\pi_i} E_i y_i^2,$$

$$\frac{1}{N^2}\left[\sum_{i \in U} F_i \dot{y}_i'^2 + \sum_{i \in U} \sum_{j \backslash \{i\} \in U} J_{ij} \dot{y}_i' \dot{y}_j' + \sum_{i \in U} E_i' y_i^2\right] \leq \frac{1}{N^2}\left[\sum_{i \in U} F_i \dot{y}_i'^2 + \sum_{i \in U} \sum_{j \backslash \{i\} \in U} J_{ij} \dot{y}_i' \dot{y}_j' + \sum_{i \in U} \frac{1}{\pi_i} E_i' y_i^2\right],$$

$$V'(\hat{\bar{Y}}_P') \leq V''(\hat{\bar{Y}}_P').$$

Next, we investigated the variance and associated estimators of the new product estimators with the MAR mechanism $\hat{\bar{Y}}_p'' = \frac{\hat{\bar{Y}}_{PL}'' \hat{\bar{X}}_{HT}}{\bar{X}} = \frac{\hat{P}''}{\bar{X}}$ both under the reverse and two- phase frameworks in Theorem 5 and Theorem 6, respectively.

Theorem 5. *Assume that a sample s of size n is selected according to UPWOR from a population U of size N. Let y and x be the study and auxiliary variables which are negatively related. Let $\hat{\bar{Y}}_p'' = \frac{\hat{\bar{Y}}_{PL}'' \hat{\bar{X}}_{HT}}{\bar{X}} = \frac{\hat{P}''}{\bar{X}}$ be the new product estimator under* the *MAR mechanism. The variance and associated estimator of $\hat{\bar{Y}}_p''$ under the reverse framework are shown as follows.*

(1) The variance of $\hat{\bar{Y}}_p''$ is,

$$V'(\hat{\bar{Y}}_p'') \cong \frac{1}{N^2}\left[\sum_{i \in U} F_i \ddot{y}_i'^2 + \sum_{i \in U} \sum_{j \backslash \{i\} \in U} J_{ij} \ddot{y}_i' \ddot{y}_j' + \sum_{i \in U} E_i'' y_i^2\right],$$

where $F_i = \frac{(1-\pi_i)}{\pi_i}$, $J_{ij} = \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}$, $E_i'' = \frac{(1-p_i)}{p_i}$, $\ddot{y}_i' = y_i + Rx_i$ and $R = \frac{\bar{Y}}{\bar{X}}$.

(2) If p is known the estimator of $V'(\hat{\bar{Y}}_p'')$ is,

$$\hat{V}_p'\left(\hat{\bar{Y}}_p''\right) \cong \frac{1}{N^2}\left[\sum_{i \in s} \hat{F}_i \hat{\ddot{y}}_i'^2 + \sum_{i \in s} \sum_{j \backslash \{i\} \in s} \hat{J}_{ij} \hat{\ddot{y}}_i' \hat{\ddot{y}}_j' + \sum_{i \in s} \frac{r_i}{p_i} \hat{E}_i'' y_i^2\right],$$

where $\hat{\ddot{y}}_i' = \frac{r_i y_i}{p_i} + \hat{R}'' x_i$, $\hat{E}_i'' = \frac{(1-p_i)}{p_i \pi_i}$, $\hat{R}'' = \hat{\bar{Y}}_p'' \big/ \hat{\bar{X}}_{HT}$ and $\hat{\bar{Y}}_{PL}'' = \frac{1}{N} \sum_{i \in s} \frac{r_i y_i}{p_i \pi_i}$.

(3) If p is unknown then the estimator of $V'\left(\hat{\bar{Y}}_p''\right)$ is

$$\hat{V}_{\hat{p}}'\left(\hat{\bar{Y}}_p''\right) \cong \frac{1}{N^2}\left[\sum_{i \in s} \hat{F}_i \hat{\hat{\ddot{y}}}_i'^2 + \sum_{i \in s} \sum_{j \backslash \{i\} \in s} \hat{J}_{ij} \hat{\hat{\ddot{y}}}_i' \hat{\hat{\ddot{y}}}_j' + \sum_{i \in s} \frac{r_i}{\hat{p}_i} \hat{\hat{E}}_i'' y_i^2\right],$$

where $\hat{\bar{Y}}''_{PL} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\hat{p}_i \pi_i}$, $\hat{\bar{X}}_{HT} = \frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}$, $\hat{\hat{R}}'' = \hat{\bar{Y}}''_{PL} \big/ \hat{\bar{X}}_{HT}$, $\hat{\hat{\ddot{y}}}'_i = \frac{r_i y_i}{\hat{p}_i} + \hat{\hat{R}}'' x_i$, $\hat{\hat{E}}''_i = \frac{(1-\hat{p}_i)}{\hat{p}_i \pi_i}$ and $\hat{p}_i$ is the estimator of p_i .

Theorem 6. *Assume that a sample s of size n is selected according to UPWOR from population U of size N . Let y and x be the study and auxiliary variables which are negatively correlated. Let $\hat{\bar{Y}}''_p = \frac{\hat{\bar{Y}}''_p \hat{\bar{X}}_{HT}}{\bar{X}} = \frac{\hat{P}''}{\bar{X}}$ be the new product estimator under the MAR mechanism. The variance and associated estimator of $\hat{\bar{Y}}''_p$ under the two-phase framework are shown as follows.*

(1) The variance of $\hat{\bar{Y}}''_p$ is,

$$V''(\hat{\bar{Y}}''_p) \cong \frac{1}{N^2}\left[\sum_{i\in U} F_i \ddot{y}_i'^2 + \sum_{i\in U}\sum_{j\setminus\{i\}\in U} J_{ij}\ddot{y}'_i\ddot{y}'_j + \sum_{i\in U}\frac{1}{\pi_i}E''_i y_i^2\right],$$

where $\ddot{y}'_i = y_i + Rx_i$, $\ddot{y}'_j = y_j + Rx_j$ and $E''_i = \frac{(1-p_i)}{p_i}$.

(2) If p_i is known then the estimator of $V''(\hat{\bar{Y}}''_p)$ is,

$$\hat{V}''_{p_i}\left(\hat{\bar{Y}}''_p\right) \cong \frac{1}{N^2}\left[\sum_{i\in s}\hat{F}_i \hat{\ddot{y}}_i'^2 + \sum_{i\in s}\sum_{j\setminus\{i\}\in s}\hat{J}_{ij}\hat{\ddot{y}}'_i\hat{\ddot{y}}'_j + \sum_{i\in s}\frac{r_i}{p_i}\frac{1}{\pi_i}\hat{E}''_i y_i^2\right],$$

where $\hat{\ddot{y}}'_i = \frac{r_i y_i}{p_i} + \hat{R}'' x_i$, $\hat{E}''_i = \frac{(1-p_i)}{p_i \pi_i}$, $\hat{R}'' = \hat{\bar{Y}}''_p \big/ \hat{\bar{X}}_{HT}$ and $\hat{\bar{Y}}''_{PL} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{p_i \pi_i}$.

(3) If p_i is unknown then the estimator of $V''(\hat{\bar{Y}}''_p)$ is

$$\hat{V}''_{p_i}\left(\hat{\bar{Y}}''_p\right) \cong \frac{1}{N^2}\left[\sum_{i\in s}\hat{F}_i \hat{\hat{\ddot{y}}}_i'^2 + \sum_{i\in s}\sum_{j\setminus\{i\}\in s}\hat{J}_{ij}\hat{\hat{\ddot{y}}}'_i\hat{\hat{\ddot{y}}}'_j + \sum_{i\in s}\frac{r_i}{\hat{p}_i}\frac{1}{\pi_i}\hat{\hat{E}}''_i y_i^2\right],$$

where $\hat{\hat{\ddot{y}}}'_i = \frac{r_i y_i}{\hat{p}_i} + \hat{\hat{R}}'' x_i$, $\hat{\hat{E}}''_i = \frac{(1-\hat{p}_i)}{\hat{p}_i \pi_i}$, $\hat{\bar{Y}}''_{PL} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\hat{p}_i \pi_i}$, $\hat{\hat{R}}'' = \hat{\bar{Y}}''_p \big/ \hat{\bar{X}}_{HT}$ and $\hat{p}_i$ is the estimator of p_i .

From Theorem 5 and Theorem 6, we see that the variance of the new product estimator under the reverse framework and two-phase framework are very similar except the terms $\sum_{i\in U}\frac{1}{\pi_i}E''_i y_i^2$ and $\sum_{i\in U}E''_i y_i^2$. Next, in Lemma 11 we showed that the variance of the new product estimator when considering under the reverse framework is always smaller than the variance under the two-phase framework.

Lemma 11. *Under the MAR mechanism. If $p_i \neq 0$ and $\pi_i \neq 0$ for all $i \in U$ then $V'(\hat{\bar{Y}}''_p) \leq V''(\hat{\bar{Y}}''_p)$.*
Some existing estimators and proposed new product estimators are shown in Table 1.

Table 1 Some existing estimators and proposed estimators

Type of Estimator	Author	Nonresponse mechanism	The estimator formula
Existing estimators	The modified Murthy's (1964)	-	$\hat{\bar{Y}}'_{PHT} = \dfrac{\frac{1}{N}\sum_{i\in s_r}\frac{y_i}{\pi_i}\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}{\bar{X}}$
	Ponkaew and Lawson (2019)	MCAR	$\hat{\bar{Y}}'_{PL} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}$
		MAR	$\hat{\bar{Y}}''_{PL} = \frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p_i}$
	Ponkaew and Lawson (2018)	MCAR	$\hat{\bar{Y}}'_{Ra} = \dfrac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p}}{\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}\bar{X}$
	Ponkaew and Lawson (2022)	MAR	$\hat{\bar{Y}}''_{Ra} = \dfrac{\frac{1}{N}\sum_{i\in s}\frac{r_i y_i}{\pi_i p_i}}{\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}\bar{X}$
Proposed estimators		MCAR	$\hat{\bar{Y}}'_{P} = \dfrac{\sum_{i\in s}\frac{r_i y_i}{\pi_i p}\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}{\bar{X}}$
		MAR	$\hat{\bar{Y}}''_{P} = \dfrac{\sum_{i\in s}\frac{r_i y_i}{\pi_i p_i}\frac{1}{N}\sum_{i\in s}\frac{x_i}{\pi_i}}{\bar{X}}$

3. Results and Discussion

3.1. Simulation studies

The simulation studies are used to compare the proposed estimators with the existing estimators. We generated data using a linear model, $y_i = \beta_0 + \beta_1 x_i + \beta_2 w_i + \beta_4 k_i + \beta_5 \varepsilon_i$ where $x_i \sim N(10,1.8)$, $w_i \sim N(30,3)$, $k_i \sim N(25,5)$ and $\varepsilon_i \sim N(0,1)$ with a population of size N = 1,000. Three levels of the correlation coefficient ρ between variables y and x consisting of $(-0.3,-0.5,-0.7)$ are considered with the 70% and 90% of response rates under both MCAR and MAR. While the correlation coefficient between variables y and k is 0.51 and the correlation coefficient between variables y and w is 0.37. Four levels of sampling fractions $f = n/N$; 5%, 10%, 30%, and 50% are considered and the samples of sizes n are drawn from a population using Midzuno's (1592) scheme. The simulation is repeated 10,000 times (M = 10,000) using the R program (R Core Team, 2021). Assume that the response probability is unknown. Under MCAR the response probability p is estimated by $\hat{p} = \sum_{i\in s}\frac{r_i}{\pi_i}\left(\sum_{i\in s}\frac{1}{\pi_i}\right)^{-1}$ while under the MAR mechanism in each level of response rate, we used the glm function in program R to estimate the value of p_i for all $i \in s$. The variables r and w with the logit link function are also used to estimate the value of p_i.

We then compared the efficiency of the proposed estimators with the existing estimators using the root mean square error defined in Definition 1.

Definition 1. *Let $\hat{\theta}$ be the estimators of a population mean estimator. Let $\theta = E(\hat{\theta})$ be the expectation of $\hat{\theta}$. We define $\hat{\theta}_{[a]}$ as the estimator in iteration m^{th} and M is the number of replications.*

The root mean square of the estimator $\hat{\theta}$ is given by $RMSE\left(\hat{\theta}\right) = \sqrt{\frac{1}{M-1}\sum_{m=1}^{M}(\hat{\theta}_{[m]} - \theta)^2}$. The results are shown in Tables 2-3.

Table 2 The root mean square error of the existing estimators and proposed estimators with 70% response rate

ρ	f	Root mean square error						
		$\hat{\hat{Y}}'_{PHT}$	$\hat{\hat{Y}}'_{PL}$	$\hat{\hat{Y}}''_{PL}$	$\hat{\hat{Y}}'_{Ra}$	$\hat{\hat{Y}}''_{Ra}$	$\hat{\hat{Y}}'_{P}$	$\hat{\hat{Y}}''_{P}$
−0.30	5%	26.55	3.90	3.64	4.20	4.09	2.53	2.24
	10%	26.28	2.44	2.12	3.41	3.20	1.73	1.48
	30%	26.19	1.91	1.53	2.29	2.03	1.04	0.74
	50%	26.16	1.78	1.36	1.97	1.68	0.84	0.48
−0.50	5%	24.34	2.39	2.18	3.40	3.26	1.97	1.66
	10%	24.04	1.62	1.58	2.26	2.12	1.51	1.22
	30%	23.98	1.00	1.29	1.75	1.69	0.93	0.63
	50%	23.89	0.99	0.85	0.94	0.74	0.73	0.41
−0.70	5%	18.66	2.25	1.80	3.33	3.19	1.71	1.58
	10%	17.47	1.40	1.36	2.21	2.07	1.30	1.12
	30%	17.41	0.90	0.84	1.26	1.06	0.82	0.58
	50%	17.34	0.85	0.76	1.15	0.99	0.71	0.38

Table 2 showed that the proposed estimators under both MCAR and MAR worked well at all levels of ρ with 70% response rate. The proposed estimator $\hat{\hat{Y}}''_{P}$ under MAR performed the best as it gave the smallest root mean square error. The higher correlation the lower root mean square error are produced. The $\hat{\hat{Y}}'_{PHT}$ performed the worst in all situations.

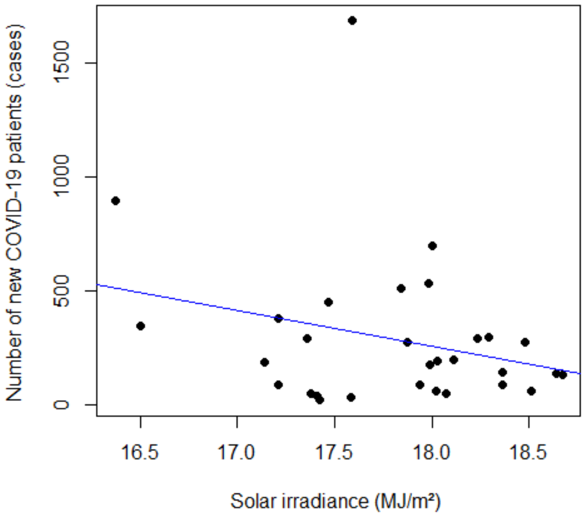
Table 3 The root mean square error of the existing estimators and proposed estimators with 90% response rate

		Root mean square error						
ρ	f	$\hat{\hat{Y}}'_{PHT}$	$\hat{\hat{Y}}'_{PL}$	$\hat{\hat{Y}}''_{PL}$	$\hat{\hat{Y}}'_{Ra}$	$\hat{\hat{Y}}''_{Ra}$	$\hat{\hat{Y}}'_{P}$	$\hat{\hat{Y}}''_{P}$
–0.30	5%	7.64	2.46	2.44	3.14	3.11	1.65	1.61
	10%	7.06	1.72	1.71	2.18	2.17	1.16	1.15
	30%	6.73	0.82	0.81	1.10	1.09	0.59	0.57
	50%	6.62	0.55	0.53	0.70	0.69	0.39	0.37
–0.50	5%	6.30	1.87	1.84	3.14	3.11	1.49	1.44
	10%	5.82	1.31	1.30	2.19	2.18	1.05	1.03
	30%	5.56	0.62	0.61	1.11	1.09	0.53	0.51
	50%	5.47	0.43	0.41	0.70	0.70	0.35	0.34
–0.70	5%	4.42	2.04	2.01	3.16	3.13	1.48	1.44
	10%	4.02	1.43	1.42	2.20	2.19	1.04	1.03
	30%	3.81	0.73	0.71	1.12	1.10	0.52	0.50
	50%	3.75	0.48	0.46	0.71	0.70	0.35	0.33

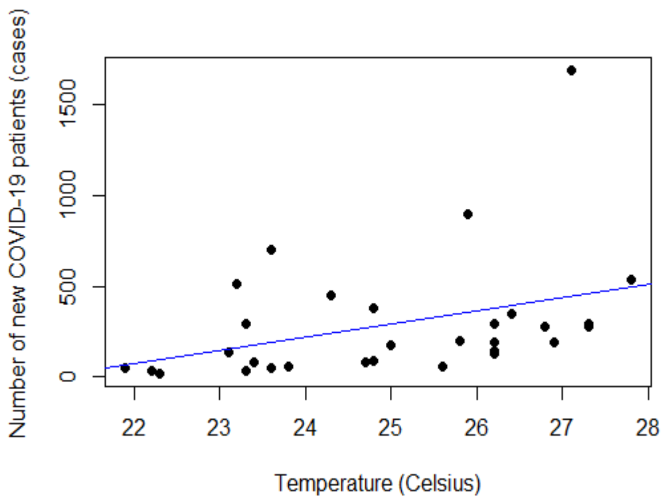
The results in Table 3 at a 90% response rate are also similar to the results shown in Table 2, a higher response rate can increase the efficiency for the estimators in terms of giving smaller errors. The proposed estimator $\hat{\hat{Y}}''_{P}$ under MAR is the best estimator in this scenario.

3.2. An application to number of COVID-19 patients in Thailand

The proposed estimators are applied to number of COVID-19 patients in Thailand 2021. The study variable y is number of new COVID-19 patients (cases) that is collected in week 52nd from 75 provinces in 2021 in Thailand from the Ministry of Public Health (2023). We use two auxiliary variables in this study, one is solar irradiance (MJ/m²) and the other one is temperature (Celsius) in Thailand. The notation of two auxiliary variables are defined by x and w respectively and they are collected by Ministry of Energy (2023). The size variable k is number of population in each province (Person) collected by National Statistical Office (2023). The variable x uses to construct the proposed estimators while the variable w is used to estimate response probability by using the logistic regression model under the MAR mechanism. The correlation coefficient between the number of new COVID-19 patients and solar irradiance is –0.27, while the correlation coefficient between the number of new COVID-19 patients and temperature is 0.37. Figures 1 and 2 show the scatter plot between the number of new COVID-19 patients and solar irradiance and temperature, respectively. We can see that from Figure 1 there is a negative correlation between the number of new COVID-19 patients and solar irradiance and from Figure 2 there is a positive correlation between the number of new COVID-19 patients and temperature.



Figures 1 A scatterplot between the number of new COVID-19 patients and solar irradiance



Figures 2 A scatterplot between the number of new COVID-19 patients and and temperature

The nonresponse rate is 4% in this study. The Midzuno's (1952) scheme is used to select a sample of size 30 records from a population of size 75 records. The results are displayed in Table 4.

Table 4 The estimated mean and variance of number of new COVID-19 patients in Thailand

Estimator	Nonresponse mechanism	Estimated mean of new COVID-19 patients	Framework	Estimated variance of mean of new COVID-19 patients
$\hat{\bar{Y}}'_{PHT}$	-	277	-	7,929.03
$\hat{\bar{Y}}'_{PL}$	MCAR	304	Reverse	3,180.19
			Two-phase	3,637.17
$\hat{\bar{Y}}''_{PL}$	MAR	302	Reverse	3,123.64
			Two-phase	3,461.16
$\hat{\bar{Y}}'_{Ra}$	MCAR	299	Reverse	3,284.89
			Two-phase	3,741.87
$\hat{\bar{Y}}''_{Ra}$	MAR	297	Reverse	3,180.94
			Two-phase	3,633.45
$\hat{\bar{Y}}'_{P}$	MCAR	308	Reverse	2,666.37
			Two-phase	3,123.35
$\hat{\bar{Y}}''_{P}$	MAR	306	Reverse	2,612.79
			Two-phase	3,021.69

The results from Table 4 also support the results found in Tables 2 and 3 which showed that the proposed product estimator $\hat{\bar{Y}}''_{P}$ under MAR performed the best with the smallest variance compared to other estimators under both frameworks. The estimated mean of new COVID-19 patients is equal to 306 cases.

4. Conclusions

New product estimators are proposed when the study variable are missing under both MCAR and MAR under UPWOR along with the variance estimators which are studied under the two-phase and reverse frameworks. We showed theoretically that the proposed estimators are almost unbiased estimators of $\bar{Y}$. when we assumed that the variable $\bar{X}$ is known. The simulation results and the application to the air pollution data showed that the proposed estimator under MAR performed the best and the proposed estimator under MCAR performed the second best in this study. In future works, we can extend the study to situations when the population mean of the auxiliary variable $\bar{X}$ is unknown and apply it to more complex survey designs. Nevertheless, the proposed estimators can assist in estimating the variable of interest when missing values occur in the real world data.

Acknowledgments

This research was funded by the National Science, Research and Innovation Fund (NSRF), and King Mongkut's University of Technology North Bangkok Contract no. KMUTNB-FF-67-B-43. We would like to thank the unknown referees for the comments.

References

Cochran WG. Sampling Techniques. New York: John Wiley & Sons; 1977.

Horvitz DG, Thompson DJ. A generalization of sampling without replacement from a finite universe. J Am Stat Assoc. 1952; 47(260): 663-685.

Midzuno H. On the sampling system with probability proportional to sum of sizes. Ann Inst Stat Math. 1952; 55(3): 99-107.

Ministry of Energy. Thailand solar irradiance data. 2023 [cited 2023 Aug 1]; Available from: http://maps.su.ac.th/solaren/.

Ministry of Plublic Health. Overview information. 2023 [cited 2023 Aug 1]; Available from: http://ddc.moph.go.th/ covid19-dashboard/?dashboard=province.

Murthy MN. Product method of estimation. Sankhya Ser A. 1964; 26(1): 69-74.

National Statistical Office. Demography population and housing branch. 2023 [cited 2023 Aug 1]; Available from: http://statbbi.nso.go.th/staticreport/page/sector/en/01.aspx.

Ponkaew C, Lawson N. A new ratio estimator for population total in the presence of nonresponse under unequal probability sampling without replacement. Thai J Math. 2018(Special Issue); 417-429.

Ponkaew C, Lawson N. Estimating variance in the presence of nonresponse under unequal probability sampling. Suranaree J Sci Tech. 2019; 26(3): 293-302.

Ponkaew C, Lawson N. New estimators for estimating population total: an application to water demand in Thailand under unequal probability sampling without replacement for missing data. Peer J. 2022; 10: e14551.

Ponkaew C, Lawson N. New variance estimators in the presence of nonresponse under unequal probability sampling, with application to fine particulate matter in Thailand. Thail Stat. 2024; 22(1): 76-101.

R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2021 [cited 2023 Sep 6]; Available from: https://www.R-project.org.