

# New estimators for estimating population total: an application to water demand in Thailand under unequal probability sampling without replacement for missing data

Chugiat Ponkaew<sup>1</sup> and Nuanpan Lawson<sup>2</sup>

<sup>1</sup> Department of Mathematics, Faculty of Science and Technology, Phetchabun Rajabhat University, Phetchabun, Thailand

<sup>2</sup> Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok, Bangsue, Bangkok, Thailand

## ABSTRACT

Water shortage could play an imperative role in the future due to an influx of water demand when compared to water supplies. Inadequate water could damage human life and other aspects related to living. This serious issue can be prevented by estimating the demand for water to bridge the small gap between demand and supplies for water. Some water consumption data recorded daily may be missing and could affect the estimated value of water demand. In this article, new ratio estimators for estimating population total are proposed under unequal probability sampling without replacement when data are missing. Two situations are considered: known or unknown mean of an auxiliary variable and missing data are missing at random for both study and auxiliary variables. The variance and associated estimators of the proposed estimators are investigated under a reverse framework. The proposed estimators are applied to data from simulation studies and empirical data on water demand in Thailand which contain some missing values, to assess the efficacies of the estimators.

Submitted 16 September 2022

Accepted 20 November 2022

Published 13 December 2022

Corresponding author

Nuanpan Lawson,  
nuanpan.n@sci.kmutnb.ac.th

Academic editor

Gowhar Meraj

Additional Information and  
Declarations can be found on  
page 19

DOI [10.7717/peerj.14551](https://doi.org/10.7717/peerj.14551)

© Copyright

2022 Ponkaew and Lawson

Distributed under

Creative Commons CC-BY 4.0

**OPEN ACCESS**

**Subjects** Natural Resource Management, Ecohydrology, Food, Water and Energy Nexus

**Keywords** Water demand, Population total, Unequal probability sampling without replacement, Taylor linearization approach, Nonlinear estimator, Logistic regression

**Mathematical Subject Classification** 62D05, 62D10

## INTRODUCTION

Increasing demand for water is highly concerning because of water supply reduction. There are many reasons that cause an increase in water demand such as the rapid growth of the human population, climate change, and so on. The world water resources consist of more water from the sea compared to available fresh water or rainwater. The amount of clean water is also affected by polluted water. Many developing countries face water scarcity and flooding issues due to climate change which can affect their sustainability in economics and lead to unsafe conditions and poor health of the population. Freshwater is used for a myriad of reasons such as household usage, business and industry, agriculture

and much more. Thailand is one of the developing countries that mainly uses freshwater in agriculture which accounts for a majority of the usage of the world's available freshwater. Metropolitan waterworks and provincial waterworks are organizations who are responsible for producing, delivering, and distributing water supply to all provinces in Thailand while also providing resources for water. The former is responsible for Bangkok, Nonthaburi, and Samut Prakan and the latter is responsible for the rest of the country. Some of the consumption of water data are missing in the database system which could lead to the wrong interpretation based on missing data. The missing data or nonresponse should be taken into consideration before processing for further analysis to make for a more powerful interpretation.

If this issue is not addressed, water shortage could lead to repercussions in the future and it would be harmful for human life because of a lack of clean water to use. The management of water resources to avoid facing water scarcity needs to be taken into consideration. Knowledge of the gap between the demand and supply of water could accommodate the strategies and policy planning for the world to be prepared for sustainable water management in order to provide sufficient water according to the demand. Estimating the water demand can benefit future planning to avoid water shortage. [Bakker et al. \(2014\)](#) investigated three models to forecast water demand in both cases with the model using weather input and not using it. The simulation results found that the model using weather input gave a maximum 11 percent of the errors which is essential in water supply system control and detecting irregularity. [Huang & Lin \(2017\)](#) proposed a system dynamics model for studying demand and supply for water resources to avoid water shortage in China. The model had been used to estimate demand and supply for Shandong in China for the next 15 years. Rainwater is also one of the sources of water usage. [Boretti & Rosa \(2019\)](#) examined the correlation between water demand, population growth, and economic growth to estimate water scarcity in the world by 2050. They found that the demand for water is growing even more than the growth of the population and economy along with a low quality of resources and water to use. [Kaewprasert, Niwitpong & Niwitpong \(2022\)](#) proposed confidence intervals for estimation of the mean of delta-gamma distribution using the Bayesian method and applied it to rainfall data in Chiang Mai Thailand.

The biased estimator, namely the ratio estimator is a popular method for estimating population total ( $Y$ ) or population mean ( $\bar{Y}$ ) of a study variable ( $y$ ) when the information of an auxiliary variable ( $x$ ) exists and is highly positively correlated with  $y$ . The ratio estimator was introduced by [Cochran \(1940\)](#) under simple random sampling without replacement (SRSWOR). The mean square error and bias of the ratio estimator are investigated by using the first order approximation of the Taylor linearization approach to transform the ratio estimator to a linear estimator. Then, the properties of the ratio estimator can be approximated from the linear estimator. [Sisodia & Dwivedi \(1981\)](#) proposed a ratio estimator when the population coefficient of variation ( $C_x$ ) of  $x$  is known. The ratio estimator when the kurtosis ( $\beta_2(x)$ ) is known was proposed by H. P. Singh & M. S. Kakran (1993, unpublished data). [Upadhyaya & Singh \(1999\)](#) suggested the ratio type estimators for estimating population mean when the  $C_x$  and  $\beta_2(x)$  are known. [Bhushan &](#)

*Kumar (2022)* suggested some classes of population mean estimators based on the optimum value of the constant to improve the efficiency of the estimators under ranked set sampling. The ratio estimators of *Cochran (1940)*, *Sisodia & Dwivedi (1981)*, H. P. Singh & M. S. Kakran (1993, unpublished data) and *Upadhyaya & Singh (1999)* require the population mean of  $x$ ,  $\bar{X}$  in order to estimate  $\bar{Y}$ . Therefore, *Perri (2004)* proposed an alternative ratio estimator namely regression-in-ratio estimator for estimating  $\bar{Y}$ . The estimator of *Perri (2004)* does not require  $\bar{X}$  by using the regression estimator to estimate this value. In other words, if the auxiliary variable  $x$  is correlated with another auxiliary variable namely  $u$  then  $\bar{X}$  can be estimated from  $u$  by using a regression estimator to estimate this value. The *Perri (2004)* estimator is a function of two estimators consisting of estimators of  $\bar{Y}$  and  $\bar{X}$ . In the context of unequal probability sampling without replacement (UPWOR), *Bacanli & Kadilar (2008)* modified the ratio estimators under SRSWOR by estimating the population mean of  $y$  and  $x$  under SRSWOR using the *Horvitz & Thompson (1952)* type estimators. The variance and associated estimators of *Bacanli & Kadilar's (2008)* estimator can be obtained by using a Taylor linearization approach and method from *Horvitz & Thompson (1952)*. *Lawson (2021)* suggested a general class of ratio estimators for population mean in the form of a combined estimator making use of known auxiliary variables such as the coefficient of variation, coefficient of skewness, coefficient of kurtosis and so on. The *Lawson (2021)* estimator performed well giving a smaller mean square error especially for a small sample size.

The ratio estimators in the full response case cannot be used to estimate population mean or population total of  $y$  when some elements in the sample units are unresponsive. *Cochran (1977)* considered the ratio estimator under SRSWOR to estimate  $\bar{Y}$  in which information on  $x$  is available for all sample units and  $\bar{X}$  is known but some elements of  $y$  in the sample units are missing.

Later, ratio estimators with their properties when nonresponse occurs in both  $y$  and  $x$  but  $\bar{X}$  is known under SRSWOR were proposed by *Rao (1986, 1987)*, *Khare & Srivastava (1997)*, *Okafor & Lee (2000)*, *Särndal & Lundström (2005)*. *Kumar (2015)*, *Lawson (2017)* introduced estimators for estimating population total and population mean and their variance estimators under probability proportional to size with replacement sampling and nonresponse present in the study. The Lawson estimators are approximately unbiased estimators and they do not require the response propensity when the response probability is uniformly nonresponse, and the sampling fraction is small. Under UPWOR and when information on  $x$  is available for all sample units when  $\bar{X}$  is known, *Ponkaew & Lawson (2018)* proposed a ratio estimator for population total of  $y$  with a uniform nonresponse. The variance and associated estimators are also discussed under a reverse framework and when the sampling fraction is ignored. In the same year, *Ponkaew (2018)* proposed a linear generalized regression estimator (GREG) for population total when information about calibration variables  $u_1, u_2, \dots, u_q$  exists. The estimator of *Ponkaew (2018)* is in a form of a nonlinear estimator then automated linearization approach was used to transform this estimator to a linear form. Consequently, the variance and their estimators can be approximated from linear estimators. The ratio estimators in the presence of nonresponse require the value of  $\bar{X}$  in both situations where nonresponse occurs with variables  $y$  and  $x$

and nonresponse occurs only with the variable  $y$ . [Ponkaew \(2018\)](#) considered the missing completely at random (MCAR) mechanism which is unlikely to occur in practice. [Lawson & Ponkaew \(2019\)](#) suggested a new GREG estimator using the idea of [Lawson \(2017\)](#) under unequal probability sampling without replacement and nonresponse occurring missing completely at random and when the sampling fraction is small and therefore can be omitted. However, their estimator requires joint inclusion probability which sometimes can be difficult to find. [Lawson & Siripanich \(2020\)](#) improved a new GREG estimator based on the idea of [Lawson & Ponkaew \(2019\)](#) for more flexible situations with the non-uniform nonresponse mechanism or missing at random (MAR) and where the sampling fractions are both large and small. [Ponkaew & Lawson \(2023\)](#) proposed a new approximately unbiased GREG estimator in the form of a ratio estimator following [Ponkaew & Lawson \(2018\)](#) and [Lawson & Ponkaew \(2019\)](#) under the same situation where nonresponse occurs under MCAR but extended it when the sampling fractions are both large and small which is in a general form.

Some researchers suggested to estimate the missing values before further analysis. For example, [Shahzad et al. \(2020\)](#) proposed population mean estimators for when there are some missing observations in the study utilizing robust regression to apply to the regression coefficient estimator under SRSWOR when outliers are present in the study. They considered when nonresponse occurs in the study variable, and in both the study and auxiliary variable when the population mean of the auxiliary variable is known and unknown. [Anas et al. \(2022\)](#) also suggested ratio type regression estimators when nonresponse is present in the study in three situations similar to [Shahzad et al. \(2020\)](#) but the quantile regression in the mean estimator when outliers are present in the study was used. [Chodjuntug & Lawson \(2022a\)](#) suggested a new imputation method to create a population mean estimator when missing data appears in the study variable and applied it to estimate fine particulate matter in Bangkok, Thailand. They suggested to apply two constants to minimize the mean square error of the population mean estimator. [Chodjuntug & Lawson \(2022b\)](#) developed a new estimator by adjusting [Chodjuntug & Lawson's \(2022a\)](#) by utilizing the response rate and the constant that minimizes the mean square error (MSE) of their proposed estimator. Their estimator using the constant that makes the minimum MSE performed the best. [Bhushan et al. \(2022\)](#) proposed some imputation methods for estimating population mean in the form of logarithmic imputations under SRSWOR for missing data.

In this article, we aim to propose new ratio estimators by extending the [Ponkaew & Lawson \(2018\)](#) estimator to situations where  $\bar{X}$  is known or unknown and nonresponse occurs with both variables  $y$  and  $x$ . In the situation where  $\bar{X}$  is unknown we used the concept from [Perri \(2004\)](#) to estimate its value from the calibration variables  $u_1, u_2, \dots, u_q$  using the linear GREG estimator of [Ponkaew \(2018\)](#). The variance and associated estimators of the proposed estimators are investigated under the reverse framework. Furthermore, the proposed ratio estimators are considered under both missing at random (MAR) which is more flexible to occur in practice and also considered under MCAR nonresponse mechanism. Finally, we compared the efficiency of the proposed estimators

and their variance estimators between the MAR and MCAR mechanisms through a simulation study and an application to water demand data in Thailand.

## MATERIALS AND METHODS

### Basic setup

In this section, we introduce notations and basic notions about the population total estimator and their variance estimators under the reverse framework. Let  $y$  be a study variable and the population total of  $y$  is  $Y = \sum_{i \in U} y_i$  where  $U = \{1, 2, \dots, N\}$  and  $N$  is the population size. Suppose the auxiliary variables  $x$ ,  $w$  and the size variable  $k$  are available and highly positively correlated with the study variable. The calibration variables  $u_1, u_2, \dots, u_q$  where  $q \geq 1$  are also available and they are correlated with the auxiliary variable  $x$ . Let,  $\mathbf{u}_i = (1 \quad u_{i1} \quad \dots \quad u_{iq})'$  and  $U_N = (\mathbf{u}_1 \quad \mathbf{u}_2 \quad \dots \quad \mathbf{u}_N)'$  be the  $N \times (q+1)$  matrix values of  $\mathbf{u}_i$ . We are using the GREG estimator model from [Särndal, Swensson & Wretman \(1992\)](#) and [Särndal \(2007\)](#) in which the linear assisting model  $\xi$ ,  $E_\xi(x_i) = \beta' \mathbf{u}_i$  and  $V_\xi(x_i) = \sigma_i^2$ . The linear assisting model  $\xi$  is a model describing the relationship between the study variable and auxiliary variable. Let  $q_i$  be determined by the linear assisting model  $\xi$  that is  $q_i = 1/\sigma_i^2$ . Usually, the standard choice of  $q_i$  is  $q_i = 1$  and it is determined by the linear assisting model  $\xi$ :  $E_\xi(x_i) = \beta' \mathbf{u}_i$  and  $V_\xi(x_i) = \sigma^2$ .

Let,  $\mathbf{F}$  be the set of all possible subsets of  $U$  and the sample  $s$  of size  $n$  was selected from the population  $U$  under UPWOR. A sampling design  $p(\cdot)$  is a probability distribution on  $\mathbf{F}$ , i.e.,  $P(s) \geq 0$  for all  $s \in \mathbf{F}$  and  $\sum_{s \in \mathbf{F}} p(s) = 1$ . Let,  $\pi_i = P(i \in s) = \sum_{s \ni i} P(s)$  be the first order inclusion probability and  $\pi_{ij} = P(i \wedge j \in s) = \sum_{s \supset \{i,j\}} P(s)$  be the second order inclusion probability. We also define  $E_S(\bullet)$  and  $V_S(\bullet)$  as the expectation and variance operators with respect to the UPWOR sampling design.

In the presence of nonresponse, let subscript  $R$  and  $r_i$  be the nonresponse mechanism and nonresponse indicator variable of  $y_i$  which  $r_i = 1$  if unit  $i$  responds to item  $y$  otherwise  $r_i = 0$ . Let,  $\mathbf{R} = (r_1 \quad r_2 \quad \dots \quad r_N)'$  be the vector of the response indicator and  $p_i = P(r_i = 1)$  be the response probability under MAR nonresponse. Let,  $E_R(\bullet)$  and  $V_R(\bullet)$  be the expectation and variance operators with respect to the nonresponse mechanism. Three assumptions are defined; (A<sub>1</sub>) the response mechanism is uniform response.

(A<sub>2</sub>)  $\hat{\beta}_r - \beta = O_p(n_r^{-\frac{1}{2}})$  and (A<sub>3</sub>)  $V\left(\sum_{i \in s} \frac{b_i}{\pi_i}\right) \rightarrow 0$  as  $n \rightarrow \infty$  where  $b_i = w_i$  or  $r_i$ . We also consider three more conditions for investigating the estimator of  $Y = \sum_{i \in U} y_i$  as follows.

(B<sub>1</sub>) nonresponse occurs only on  $y$ , the information on  $x_i$  is available for all  $i \in s$  and  $\bar{X}$  is known. (B<sub>2</sub>) nonresponse occurs on both  $y$  and  $x$  and  $\bar{X}$  is known and (B<sub>3</sub>) nonresponse occurs both with  $y$  and  $x$  and  $\bar{X}$  is unknown but information on  $u_1, u_2, \dots, u_q$  are available for all  $i \in s$  and  $\bar{U}_j = \frac{1}{N} \sum_{i \in U} u_{ij}$ ,  $j = 1, 2, \dots, q$  are known.

Throughout this article, we consider variance estimation of the population total estimator in the presence of nonresponse under the reverse framework. Therefore, we discuss three steps to investigate the variance and its nonlinear estimator such as the ratio estimator when nonresponse occurs in the study variable. Assume that we have  $K$  variables consisting of  $t_1$ , the study variable and  $t_2, t_3, \dots, t_K$ , auxiliary variables. Let  $\hat{Y}_s$  be a nonlinear estimator and be defined by,

$$\hat{Y}_s = \psi(\hat{\mathbf{T}}), \quad (1)$$

where  $\psi$  is a known smooth function,  $\hat{\mathbf{T}} = [\hat{t}_1 \quad \hat{t}_2 \quad \dots \quad \hat{t}_K]$ ,  $K \geq 2$   $\hat{t}_k = \sum_{i \in S} \frac{r_i t_{ki}}{\pi_i p_i}$  if the variable  $t_k$  exhibits nonresponse otherwise it can be obtained by  $\hat{Y}_s$ . Under the reverse framework, variance of  $\hat{t}_k = \sum_{i \in S} \frac{t_{ki}}{\pi_i}$

$$V(\hat{Y}_s) = E_R V_S(\hat{Y}_s | \mathbf{R}) + V_R E_S(\hat{Y}_s | \mathbf{R}) = V_1 + V_2, \quad (2)$$

where  $V_1 = E_R V_S(\hat{Y}_s | \mathbf{R})$  and  $V_2 = V_R E_S(\hat{Y}_s | \mathbf{R})$ . The formula of  $V(\hat{Y}_s)$  consists of three steps as below.

**Step 1:** Investigate a formula of  $V_1 = E_R V_S(\hat{Y}_s | \mathbf{R})$ . Since  $\hat{Y}_s$  is in a form of a nonlinear estimator then  $V_1 = E_R V_S(\hat{Y}_s | \mathbf{R})$  can be approximated by,

$$V'_1 \cong E_R V_S(\hat{Y}_{s(1)} | \mathbf{R}), \quad (3)$$

where  $\hat{Y}_{s(1)}$  is a linear estimator of  $\hat{Y}_s$  under the Taylor linearization approach.

**Step 2:** Investigate the formula of  $V_2 = V_R E_S(\hat{Y}_s | \mathbf{R})$ . The formula of  $V_2 = V_R E_S(\hat{Y}_s | \mathbf{R})$  can be approximated by,

$$V'_2 \cong V_R(Y_s^{\%} | \mathbf{R}), \quad (4)$$

where  $Y_s^{\%} = E_S(\hat{Y}_s | \mathbf{R})$ .

**Step 3:** Approximate the value of  $V(\hat{Y}_s)$  and its estimator. The value of  $V(\hat{Y}_s)$  can be obtained by,

$$V(\hat{Y}_s) = V'_1 + V'_2. \quad (5)$$

The estimator of  $V(\hat{Y}_s)$  can be obtained by substituting estimators for the unknown parameter in (5). Then, the estimator of  $V(\hat{Y}_s)$  is defined by,

$$\hat{V}(\hat{Y}_s) = \hat{V}'_1 + \hat{V}'_2, \quad (6)$$

where  $\hat{V}'_1, \hat{V}'_2$  are the estimators of  $V'_1, V'_2$  respectively.

### Existing estimators under uniform nonresponse

Uniform nonresponse or missing completely at random (MCAR) is a nonresponse mechanism in which the probability of response of the study variable  $y$  neither depends on itself nor another variable such as  $x, k$  or  $w$ . In this section, we discuss two estimators for estimating population total in the presence of uniform nonresponse namely ratio and GREG estimators proposed by [Ponkaew & Lawson \(2018\)](#) and [Ponkaew \(2018\)](#), respectively. The variance estimation of both ratio and GREG estimators are considered

under the reverse framework and the sampling fraction is negligible with the UPWOR sampling design.

### The ratio estimator

When nonresponse occurs only with  $y$  but the population mean and its estimator of  $x$  are available, [Ponkaew & Lawson \(2018\)](#) proposed ratio estimators to estimate population mean and the total of  $y$  under unequal probability sampling without replacement and the nonresponse mechanism is MCAR. The [Ponkaew & Lawson \(2018\)](#) estimator for population mean is

$$\hat{Y}_R = \frac{\frac{1}{N} \sum_{i \in s} \frac{r_i y_i}{\pi_i p}}{\frac{1}{N} \sum_{i \in s} \frac{x_i}{\pi_i}} \bar{X} = \frac{\hat{Y}_r}{\hat{X}_{HT}} \bar{X} = \hat{R}_r \bar{X}, \quad (7)$$

where  $\hat{Y}_r = \frac{1}{N} \sum_{i \in s} \frac{r_i y_i}{\pi_i p}$ ,  $\hat{X}_{HT} = \frac{1}{N} \sum_{i \in s} \frac{x_i}{\pi_i}$  and  $\hat{R}_r = \hat{Y}_r (\hat{X}_{HT})^{-1}$ . [Ponkaew & Lawson's \(2018\)](#) estimator for population total is

$$\hat{Y}_R = N \hat{Y}_R = N \bar{X} \hat{R}_r. \quad (8)$$

We note that, if  $p$  is unknown the estimator of  $p$  is equal to  $\hat{p} = \left( \sum_{i \in s} \frac{r_i}{\pi_i} \right) \left( \sum_{i \in s} \frac{1}{\pi_i} \right)^{-1}$ .

The variance and associated estimators of the estimator in (8) is defined in (9),

$$V(\hat{Y}_R) = \sum_{i \in U} D_i (y_i - R x_i)^2 + \sum_{i \in U} \sum_{i \setminus \{j\} \in U} D_{ij} (y_i - R x_i)(y_j - R x_j) + \sum_{i \in U} \frac{1 - p_i}{p_i} y_i^2, \quad (9)$$

where  $R = \bar{Y} \bar{X}^{-1}$ . The estimator of  $V(\hat{Y}_R)$  is given in (10),

$$\hat{V}(\hat{Y}_R) = \sum_{i \in s} \hat{D}_i (y_i - \hat{R}_r x_i)^2 + \sum_{i \in s} \sum_{i \setminus \{j\} \in s} \hat{D}_{ij} (y_i - \hat{R}_r x_i)(y_j - \hat{R}_r x_j) + \sum_{i \in s} \hat{E}_i y_i^2, \quad (10)$$

where  $\hat{R}_r = \sum_{i \in s} \frac{r_i y_i}{\pi_i p} \left( \sum_{i \in s} \frac{x_i}{\pi_i} \right)^{-1}$ ,  $\hat{E}_i = \frac{r_i E_i}{\pi_i}$ ,  $\hat{D}_i = \frac{r_i D_i}{p}$  and  $\hat{D}_{ij} = \frac{r_i r_j D_{ij}}{p^2}$ .

### The GREG estimator

The GREG estimators for estimating population mean or population total of the study variable is a powerful method when the calibration variables  $u_1, u_2, \dots, u_q$  are present where  $q \geq 1$  are also available. In full response, [Särndal, Swensson & Wretman \(1992\)](#) and [Särndal \(2007\)](#) proposed a GREG estimator under the linear assisting model  $\xi$ ,

$$E_\xi(x_i) = \beta' u_i \text{ and } V_\xi(x_i) = \sigma_i^2. \quad (11)$$

Let  $Q_s = \text{diag}(q_i)_{s \times s}$  and  $q_i$  be determined by the linear assisting model  $\xi$  in (5.1) i.e.,  $q_i = \sigma_i^{-2}$ . In the presence of nonresponse, [Särndal & Lundström \(2005\)](#) proposed a linear GREG estimator to estimate population total. They investigated variance and associated estimators under the two-phase framework. [Ponkaew \(2018\)](#) proposed linear GREG estimators for estimating the population mean of  $x$  under the MCAR mechanism which is defined by,

$$\begin{aligned}\hat{X}_{GREG}^{(1)} &= \frac{1}{N} \sum_{i \in s} \frac{r_i x_i}{\pi_i p} + \left[ \bar{U} - \frac{1}{N} \sum_{i \in s} \frac{r_i \mathbf{u}_i}{\pi_i p} \right]' \left( \sum_{i \in s} \frac{r_i q_i \mathbf{u}_i \mathbf{u}'_i}{\pi_i p} \right)^{-1} \left( \sum_{i \in s} \frac{r_i q_i \mathbf{u}_i x_i}{\pi_i p} \right) \\ &= \hat{X}_r^{(1)} + \left[ \bar{U} - \hat{U}_r^{(1)} \right]' \hat{\beta}_r,\end{aligned}\quad (12)$$

where  $\hat{X}_r^{(1)} = \frac{1}{N} \sum_{i \in s} \frac{r_i x_i}{\pi_i p}$ ,  $\hat{U}_r^{(1)} = \frac{1}{N} \sum_{i \in s} \frac{r_i \mathbf{u}_i}{\pi_i p}$ , and

$$\hat{\beta}_r = \left( \sum_{i \in s} \frac{r_i q_i \mathbf{u}_i \mathbf{u}'_i}{\pi_i p} \right)^{-1} \left( \sum_{i \in s} \frac{r_i q_i \mathbf{u}_i x_i}{\pi_i p} \right) = \left( \sum_{i \in s} \frac{r_i q_i \mathbf{u}_i \mathbf{u}'_i}{\pi_i} \right)^{-1} \left( \sum_{i \in s} \frac{r_i q_i \mathbf{u}_i x_i}{\pi_i} \right).$$

Then, the GREG estimator to estimate the population total of  $x$  is

$$\begin{aligned}\hat{X}_{GREG}^{(1)} &= N \hat{X}_{GREG}^{(1)} = \sum_{i \in s} \frac{r_i x_i}{\pi_i p} + \left[ U - \sum_{i \in s} \frac{r_i \mathbf{u}_i}{\pi_i p} \right]' \left( \sum_{i \in s} \frac{r_i q_i \mathbf{u}_i \mathbf{u}'_i}{\pi_i p} \right)^{-1} \left( \sum_{i \in s} \frac{r_i q_i \mathbf{u}_i x_i}{\pi_i p} \right), \\ &= \hat{X}_r^{(1)} + \left[ U - \hat{U}_r^{(1)} \right]' \hat{\beta}_r,\end{aligned}\quad (13)$$

where  $\hat{X}_r^{(1)} = N \hat{X}_r^{(1)} = \sum_{i \in s} \frac{r_i x_i}{\pi_i p}$ ,  $\hat{U}_r^{(1)} = N \hat{U}_r^{(1)} = \sum_{i \in s} \frac{r_i \mathbf{u}_i}{\pi_i p}$ .

Under the reverse framework and when sampling fraction is negligible the variance of  $\hat{W}_{GREG}^{(1)}$  is

$$V(\hat{X}_{GREG}^{(1)}) \cong \sum_{i \in U} D_{1i} e_i^2 + \sum_{i \in U} \sum_{j \in U \setminus \{i\}} D_{ij} e_i e_j, \quad (14)$$

where  $D_{1i} = \frac{(1 - \pi_i)}{\pi_i p}$ ,  $D_{ij} = \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}$  and  $e_i = (x_i - \mathbf{u}'_i \beta)$ .

The estimator of  $V(\hat{W}_{GREG}^{(1)})$  is equal to.

$$\hat{V}(\hat{X}_{GREG}^{(1)}) \cong \frac{1}{p'^2} \left[ \sum_{i \in s} \hat{D}_i r_i \hat{e}_i^2 + \sum_{i \in s} \sum_{j \in s \setminus \{i\}} \hat{D}_{ij} r_i r_j \hat{e}_i \hat{e}_j \right], \quad (15)$$

where  $\hat{e}_i = (x_i - \mathbf{u}'_i \hat{\beta}_r)$ ,  $\hat{D}_i = \frac{1 - \pi_i}{\pi_i^2}$ ,  $\hat{D}_{ij} = \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij} \pi_i \pi_j}$ ,  $p' = \left( \sum_{i \in s} \frac{1}{\pi_i} \right) \left( \sum_{i \in s} \frac{r_i}{\pi_i} \right)^{-1}$  if  $p$  is unknown otherwise  $p' = p$ .

## RESULTS AND DISCUSSION

### The proposed new ratio estimators

In the previous section, we introduced two estimators of the population total: ratio and GREG estimators in the presence of uniform nonresponse. The variance estimation for both ratio and GREG estimators are considered under the UPWOR sampling design and when the sampling fraction is negligible. However, the ratio estimators in (7) and (8) are considered under a situation where nonresponse occurs in  $y$  only and they require the



value of the population mean of  $x$ . Then, in this section we aim to propose new ratio estimators when nonresponse occurs in both variables  $y$  and  $x$ . We also consider two distinct situations of  $\bar{X}$  that are known or unknown. In the situation where  $\bar{X}$  is unknown we estimate it from the calibration variables  $u_1, u_2, \dots, u_q$  using the GREG estimator. In the context of nonresponse, we investigate the proposed ratio estimator under the MAR mechanism because it has weak assumptions and tends to occur in real life more often than the MCAR mechanism. However, we still consider new ratio estimators under the MCAR mechanism for comparing the efficiency of the proposed estimators. First of all, we extended the [Ponkaew & Lawson \(2018\)](#) estimators to the MAR mechanism. The ratio estimator of [Ponkaew & Lawson \(2018\)](#) for estimating population mean under the MAR mechanism is equal to,

$$\hat{Y}_R^{(1)} = \frac{\frac{1}{N} \sum_{i \in s} \frac{r_i y_i}{\pi_i p_i}}{\frac{1}{N} \sum_{i \in s} \frac{x_i}{\pi_i}} \bar{X} = \frac{\hat{Y}_r^{(1)}}{\hat{X}_{HT}} \bar{X} = \hat{R}_r^{(1)} \bar{X}, \quad (16)$$

$$\text{where } \hat{Y}_r^{(1)} = \frac{1}{N} \sum_{i \in s} \frac{r_i y_i}{\pi_i p_i}, \hat{X}_{HT} = \frac{1}{N} \sum_{i \in s} \frac{x_i}{\pi_i} \text{ and } \hat{R}_r^{(1)} = \hat{Y}_r^{(1)} (\hat{X}_{HT})^{-1}.$$

Then, the ratio estimator for estimating population total under the MAR mechanism is

$$\hat{Y}_R^{(1)} = N \hat{Y}_R^{(1)} = N \bar{X} \hat{R}_r^{(1)}, \quad (17)$$

Under the MAR mechanism if  $p_i$  is unknown then it is estimated using the probit or logistic regression models. The variance and associated estimators of  $\hat{Y}_R^{(1)}$  are discussed in Theorem 4.1.

**Theorem 1.** Under condition  $(B_1)$  with the reverse framework and the nonresponse mechanism is MAR.

(1) The variance of  $\hat{Y}_R^{(1)}$  is

$$V\left(\hat{Y}_R^{(1)}\right) \cong \sum_{i \in U} D_i A_i^2 + \sum_{i \in U} \sum_{i \setminus \{j\} \in U} D_{ij} A_i A_j + \sum_{i \in U} E_i y_i^2,$$

$$\text{where } A_i = y_i - R x_i, R = \bar{Y} \bar{X}^{-1} \text{ and } E_i = \frac{1 - p_i}{p_i}.$$

(2) The estimator of  $V\left(\hat{Y}_R^{(1)}\right)$  is

$$\hat{V}\left(\hat{Y}_R^{(1)}\right) \cong \sum_{i \in s} \hat{D}_i \hat{A}_i^{(1)2} + \sum_{i \in s} \sum_{i \setminus \{j\} \in s} \hat{D}_{ij} \hat{A}_i^{(1)} \hat{A}_j^{(1)} + \sum_{i \in s} \hat{E}_i y_i^2,$$

$$\text{where } \hat{A}_i^{(1)} = y_i - \hat{R}_r^{(1)} x_i, \hat{R}_r^{(1)} = \sum_{i \in s} \frac{r_i y_i}{\pi_i p_i} \left( \sum_{i \in s} \frac{x_i}{\pi_i} \right)^{-1}, \hat{E}_i = \frac{r_i E_i}{\pi_i}, \hat{D}_i = \frac{r_i D_i}{p_i} \text{ and } \hat{D}_{ij} = \frac{r_i r_j D_{ij}}{p_i p_j}.$$

**Proof.** Let  $\hat{Y}_R^{(1)}$  be defined in (17). Therefore, variance of  $\hat{Y}_R^{(1)}$  is

$$V\left(\hat{Y}_R^{(1)}\right) = V\left(N \bar{X} \hat{R}_r^{(1)}\right) = N^2 \bar{X}^2 V\left(\hat{R}_r^{(1)}\right). \quad (18)$$

Furthermore, the estimator of  $V\left(\hat{Y}_R^{(1)}\right)$  can be obtained by,

$$\hat{V}(\hat{Y}_R^{(1)}) = N^2 \bar{X}^2 \hat{V}(\hat{R}_r^{(1)}). \quad (19)$$

Since  $\hat{R}_r^{(1)}$  is a nonlinear estimator then the variance of this estimator is equal to,

$$V(\hat{R}_r^{(1)}) = E_R V_S(\hat{R}_r^{(1)} | \mathbf{R}) + V_R E_S(\hat{R}_r^{(1)} | \mathbf{R}) = V_1 + V_2, \quad (20)$$

where  $V_1 = E_R V_S(\hat{R}_r^{(1)} | \mathbf{R})$ ,  $V_2 = V_R E_S(\hat{R}_r^{(1)} | \mathbf{R})$ .

**Step 1:** Investigate the formula of  $V_1 = E_R V_S(\hat{R}_r^{(1)} | \mathbf{R})$ .

By using the Taylor linearization approach the linear estimator of  $\hat{R}_r^{(1)}$  is

$$\hat{R}_r^{(1)} \equiv \text{Constant} + \frac{1}{N\bar{X}} \sum_{i \in S} \frac{\tilde{A}_i}{\pi_i}, \quad (21)$$

where  $\tilde{A}_i = \left( \frac{r_i y_i}{p_i} - \tilde{R}_r^{(1)} x_i \right)$ . Then  $V_1 = E_R V_S(\hat{R}_r^{(1)} | \mathbf{R})$  can be approximated by,

$$\begin{aligned} V'_1 &\cong E_R V_S(\hat{R}_r^{(1)} | \mathbf{R}) = E_R V_S \left( \text{Constant} + \frac{1}{N\bar{X}} \sum_{i \in S} \frac{\tilde{A}_i}{\pi_i} \middle| \mathbf{R} \right) \\ &= \frac{1}{N^2 \bar{X}^2} E_R \left( \sum_{i \in U} D_i \tilde{A}_i^2 + \sum_{i \in U} \sum_{i \setminus \{j\} \in U} D_{ij} \tilde{A}_i \tilde{A}_j \middle| \mathbf{R} \right) \\ &= \frac{1}{N^2 \bar{X}^2} \left( \sum_{i \in U} D_i A_i^2 + \sum_{i \in U} \sum_{i \setminus \{j\} \in U} D_{ij} A_i A_j \right), \end{aligned}$$

where  $A_i = E_R V_S(\tilde{A}_i | \mathbf{R}) = y_i - R x_i$  and  $R = \bar{Y} \bar{X}^{-1}$ .

Therefore,

$$V'_1 \cong \frac{1}{N^2 \bar{X}^2} \left( \sum_{i \in U} D_i A_i^2 + \sum_{i \in U} \sum_{i \setminus \{j\} \in U} D_{ij} A_i A_j \right). \quad (22)$$

**Step 2:** Investigate the formula of  $V_2 = V_R E_S(\hat{R}_r^{(1)} | \mathbf{R})$ .

The formula of  $V_2 = V_R E_S(\hat{R}_r^{(1)} | \mathbf{R})$  can be approximated by,

$$\begin{aligned} V'_2 &\cong V_R E_R(\hat{R}_r^{(1)} | \mathbf{R}) = V_R E_R \left( \frac{\frac{1}{N} \sum_{i \in S} \frac{r_i y_i}{\pi_i p_i}}{\frac{1}{N} \sum_{i \in S} \frac{x_i}{\pi_i}} \middle| \mathbf{R} \right) = V_R \left( \frac{\frac{1}{N} \sum_{i \in U} \frac{r_i y_i}{p_i}}{\bar{X}} \middle| \mathbf{R} \right) \\ &= \frac{1}{N^2 \bar{X}^2} \sum_{i \in U} \frac{(1 - p_i) y_i^2}{p_i} = \frac{1}{N^2 \bar{X}^2} \sum_{i \in U} E_i y_i^2 \end{aligned}$$

where  $E_i = \frac{(1 - p_i)}{p_i}$ .  
Then,

$$V'_2 \cong \frac{1}{N^2 \bar{X}^2} \sum_{i \in U} E_i y_i^2. \quad (23)$$

**Step 3:** Approximate the value of  $V(\hat{R}_r^{(1)})$  and its estimators.

The value of  $V(\hat{R}_r^{(1)})$  can be approximated by,

$$V(\hat{R}_r^{(1)}) \cong V'_1 + V'_2 = \frac{1}{N^2 \bar{X}^2} \left( \sum_{i \in U} D_i A_i^2 + \sum_{i \in U} \sum_{i \setminus \{j\} \in U} D_{ij} A_i A_j + \sum_{i \in U} E_i y_i^2 \right). \quad (24)$$

The estimator of  $V(\hat{R}_r^{(1)})$  is

$$\hat{V}(\hat{R}_r^{(1)}) = \frac{1}{N^2 \bar{X}^2} \left( \sum_{i \in s} \hat{D}_i \hat{A}_i^{(1)2} + \sum_{i \in s} \sum_{i \setminus \{j\} \in s} \hat{D}_{ij} \hat{A}_i^{(1)} \hat{A}_j^{(1)} + \sum_{i \in s} \hat{E}_i y_i^2 \right). \quad (25)$$

Replace (25) into (18) then the variance of  $\hat{Y}_R^{(1)}$  is

$$V(\hat{Y}_R^{(1)}) \cong \sum_{i \in U} D_i A_i^2 + \sum_{i \in U} \sum_{i \setminus \{j\} \in U} D_{ij} A_i A_j + \sum_{i \in U} E_i y_i^2. \quad (26)$$

Furthermore, the estimator of  $V(\hat{Y}_R^{(1)})$  can be obtained by substituting (26) in (19) then,

$$\hat{V}(\hat{Y}_R^{(1)}) \cong \sum_{i \in s} \hat{D}_i \hat{A}_i^{(1)2} + \sum_{i \in s} \sum_{i \setminus \{j\} \in s} \hat{D}_{ij} \hat{A}_i^{(1)} \hat{A}_j^{(1)} + \sum_{i \in s} \hat{E}_i y_i^2. \quad (27)$$

In (16) and (17), we extend the ratio estimators of [Ponkaew & Lawson \(2018\)](#) to the MAR mechanism and discussed the variance and its estimators in Theorem 1. However, the ratio estimator for population mean in (16) and for population total in (17) can be used under the condition ( $B_1$ ) that is, when nonresponse occurs only with the  $y$  variable but information on  $x_i$  for all  $i \in s$  and  $\bar{X}$  needs to be known. Next, we proposed new ratio estimators under condition ( $B_2$ ) where nonresponse occurs on both  $y$  and  $x$  but  $\bar{X}$  is known and condition ( $B_3$ ) nonresponse occurs both  $y$  and  $x$  and  $\bar{X}$  is unknown but information of  $u_1, u_2, \dots, u_q$  are available for all  $i \in s$  and the population mean of  $u_1, u_2, \dots, u_q$  are also known.

### The new ratio estimator when $\bar{X}$ is known

Assume that the condition ( $B_2$ ) is satisfied when nonresponse occurs with both variables  $y$  and  $x$  but  $\bar{X}$  is known. The new ratio estimator for estimating population mean is given below,

$$\hat{Y}_R^{(2)} = \left( \frac{\frac{1}{N} \sum_{i \in s} \frac{r_i y_i}{\pi_i p_i}}{\frac{1}{N} \sum_{i \in s} \frac{r_i x_i}{\pi_i p_i}} \right) \frac{1}{N} \sum_{i \in U} x_i = \frac{\hat{Y}_r}{\hat{X}_r} \bar{X} = \hat{R}_r^{(2)} \bar{X}, \quad (28)$$

where  $\hat{Y}_r = \frac{1}{N} \sum_{i \in S} \frac{r_i y_i}{\pi_i p_i}$ ,  $\hat{X}_r = \frac{1}{N} \sum_{i \in S} \frac{r_i x_i}{\pi_i p_i}$ ,  $\hat{R}_r^{(2)} = \hat{Y}_r (\hat{X}_r)^{-1}$ . Furthermore, the estimator of  $p_i$  can be obtained by using the probit or logistic regression models under the MAR mechanism. Then, the new ratio estimator for the population total is

$$\hat{Y}_R^{(2)} = N \hat{Y}_R^{(2)} = N \bar{X} \hat{R}_r^{(2)}. \quad (29)$$

The variance and associated estimators of  $\hat{Y}_R^{(2)}$  are discussed in Theorem 2.

**Theorem 2.** Under condition (B<sub>2</sub>) with reverse framework and where the nonresponse mechanism is MAR.

(1) The variance of  $\hat{Y}_R^{(2)}$  is

$$V\left(\hat{Y}_R^{(2)}\right) = \sum_{i \in U} D_i A_i^2 + \sum_{i \in U} \sum_{i \setminus \{j\} \in U} D_{ij} A_i A_j + \sum_{i \in U} E_i y_i^2, \quad (30)$$

where  $A_i = y_i - R x_i$ ,  $R = \bar{Y} \bar{X}^{-1}$  and  $E_i = \frac{1 - p_i}{p_i}$ .

(2) The estimator of  $V\left(\hat{Y}_R^{(2)}\right)$  is

$$\hat{V}\left(\hat{Y}_R^{(2)}\right) = \sum_{i \in S} \hat{D}_i \hat{A}_i^{(2)2} + \sum_{i \in S} \sum_{i \setminus \{j\} \in S} \hat{D}_{ij} \hat{A}_i^{(2)} \hat{A}_j^{(2)} + \sum_{i \in S} \hat{E}_i y_i^2, \quad (31)$$

where  $\hat{A}_i^{(2)} = y_i - \hat{R}_r^{(2)} x_i$ ,  $\hat{R}_r^{(2)} = \sum_{i \in S} \frac{r_i y_i}{\pi_i p'_i} \left( \sum_{i \in S} \frac{r_i x_i}{\pi_i p'_i} \right)^{-1}$ ,  $\hat{D}_i = \frac{r_i D_i}{\pi_i p'_i}$ ,  $\hat{D}_{ij} = \frac{r_i D_{ij}}{\pi_i p'_i}$ ,  $\hat{E}_i = \frac{r_i E_i}{\pi_i p'_i}$ .

The value of  $p'_i$ ,  $p'_i = p_i$  if  $p_i$  is known otherwise  $p'_i = \hat{p}_i$ .  $\hat{p}_i$  is the estimator of  $p_i$  from the probit or logistic regression models.

The proof in Theorem 2 is similar to the proof in Theorem 1.

In Theorem 2 we investigated the variance and its estimators of  $\hat{Y}_R^{(2)}$ . We note that the variance formulas  $\hat{Y}_R^{(1)}$  and  $\hat{Y}_R^{(2)}$  are the same but the variance estimators of  $\hat{Y}_R^{(1)}$  and  $\hat{Y}_R^{(2)}$  are slightly different because the estimators of  $A_i = y_i - R x_i$  are different.

In (28) and (29) we proposed new ratio estimators for population mean and population total of the study variable when nonresponse occurs on both  $y$  and  $x$  variables but  $\bar{X}$  is known under the MAR mechanism. Furthermore, the variance and its estimator are also discussed in Theorem 2. Next, we proposed the special case of  $\hat{Y}_R^{(2)}$  when the response probability is consider under the MCAR mechanism ( $p_i = p$  for all  $i \in U$ ). Under the MAR mechanism the population mean estimator is equal to

$$\hat{Y}_R^{(2)} = \left( \frac{\frac{1}{N} \sum_{i \in S} \frac{r_i y_i}{\pi_i p}}{\frac{1}{N} \sum_{i \in S} \frac{r_i x_i}{\pi_i p}} \right) \frac{1}{N} \sum_{i \in U} x_i = \left( \frac{\frac{1}{N} \sum_{i \in S} \frac{r_i y_i}{\pi_i}}{\frac{1}{N} \sum_{i \in S} \frac{r_i x_i}{\pi_i}} \right) \frac{1}{N} \sum_{i \in U} x_i = \frac{\hat{Y}_r'}{\hat{X}_r'} \bar{X} = \hat{R}_r^{(2)} \bar{X}, \quad (32)$$

where  $\hat{Y}_r' = \frac{1}{N} \sum_{i \in S} \frac{r_i y_i}{\pi_i}$ ,  $\hat{X}_r' = \frac{1}{N} \sum_{i \in S} \frac{r_i x_i}{\pi_i}$ ,  $\hat{R}_r^{(2)} = \hat{Y}_r' (\hat{X}_r')^{-1}$ . Then, the population total estimator is

$$\hat{Y}_R^{''(2)} = N\hat{Y}_R^{''(2)} = N\bar{X}\hat{R}_r^{''(2)}. \quad (33)$$

Finally, the variance and associated estimators of  $y$  are discussed in Lemma 3.

**Lemma 3.** Under condition  $(B_2)$  with reverse framework and where the nonresponse mechanism is MCAR.

(1) The variance of  $\hat{Y}_R^{''(2)}$  is

$$V\left(\hat{Y}_R^{''(2)}\right) = \sum_{i \in U} D_i A_i^2 + \sum_{i \in U} \sum_{i \setminus \{j\} \in U} D_{ij} A_i A_j + \sum_{i \in U} E'_i y_i^2, \quad (34)$$

where  $A_i = y_i - R x_i$ ,  $R = \bar{Y}\bar{X}^{-1}$  and  $E'_i = \frac{1-p}{p}$ .

(2) The estimator of  $V\left(\hat{Y}_R^{''(2)}\right)$  is

$$\hat{V}\left(\hat{Y}_R^{''(2)}\right) = \sum_{i \in s} \hat{D}_i \hat{A}_i^{(2)2} + \sum_{i \in s} \sum_{i \setminus \{j\} \in s} \hat{D}_{ij} \hat{A}_i^{(2)} \hat{A}_j^{(2)} + \sum_{i \in s} \hat{E}'_i y_i^2, \quad (35)$$

where  $\hat{A}_i^{(2)} = y_i - \hat{R}_r^{(2)} x_i$ ,  $\hat{R}_r^{(2)} = \sum_{i \in s} \frac{r_i y_i}{\pi_i} \left( \sum_{i \in s} \frac{r_i x_i}{\pi_i} \right)^{-1}$ ,  $\hat{D}_i = \frac{r_i D_i}{\pi_i p'}$ ,  $\hat{D}_{ij} = \frac{r_i D_{ij}}{\pi_i p'}$ ,  $\hat{E}'_i = \frac{r_i E_i}{\pi_i p'}$ .

The value of  $p'$ ,  $p' = p$  if  $p$  is known otherwise  $p = \hat{p}$ .  $\hat{p}$  is the estimator of  $p$  under the

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

### The new ratio estimator when $\bar{X}$ is unknown

Assume that the condition  $(B_3)$  is satisfied,  $\bar{X}$  is unknown and nonresponse occurs on both  $y$  and  $x$  variables. However, the information of variable  $u_1, u_2, \dots, u_q$  is available for all  $i \in s$  and  $\bar{U}$  is known. Furthermore, variables  $u_1, u_2, \dots, u_q$  are highly correlated with  $x$ . Then, we extended the GREG estimator of [Ponkaew's \(2018\)](#) to the MAR mechanism and it is defined by

$$\hat{X}_{GREG} = \hat{X}_r + \left[ \bar{U} - \hat{U}_r \right]' \hat{\beta}_r, \quad (36)$$

where  $\hat{X}_r = \frac{1}{N} \sum_{i \in s} \frac{r_i x_i}{\pi_i p_i}$ ,  $\hat{U}_r = \frac{1}{N} \sum_{i \in s} \frac{r_i u_i}{\pi_i p'_i}$ ,  $\hat{U} = \frac{1}{N} \sum_{i \in U} u_i$ ,

$$\hat{\beta}_r = \left( \sum_{i \in s} \frac{r_i q_i u_i u'_i}{\pi_i p_i} \right)^{-1} \left( \sum_{i \in s} \frac{r_i q_i u_i w_i}{\pi_i p_i} \right).$$

The new ratio estimator for population mean is

$$\hat{Y}_R^{(3)} = \frac{\frac{1}{N} \sum_{i \in s} \frac{r_i y_i}{\pi_i p_i}}{\frac{1}{N} \sum_{i \in s} \frac{r_i x_i}{\pi_i p_i}} \hat{X}_{GREG}. \quad (37)$$

Then, the new ratio estimator for population total is

$$\hat{Y}_R^{(3)} = N\hat{Y}_R^{\prime(3)}. \quad (38)$$

The variance and associated estimators of  $\hat{Y}_R^{\prime(3)}$  are discussed in Theorem 4.

**Theorem 4.** Under condition  $(B_3)$  with reverse framework and nonresponse mechanism is MAR.

(1) The variance of  $\hat{Y}_R^{\prime(3)}$  is

$$V\left(\hat{Y}_R^{\prime(3)}\right) = \sum_{i \in U} D_i B_i^2 + \sum_{i \in U} \sum_{i \setminus \{j\} \in U} D_{ij} B_i B_j + \sum_{i \in U} E_i B_i^2, \quad (39)$$

where  $D_i = \frac{1 - \pi_i}{\pi_i}$ ,  $D_{ij} = \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}$ ,  $E_i = \frac{1 - p_i}{p_i}$  and  $B_i = x_i - \mathbf{R}\mathbf{u}'_i \boldsymbol{\beta}$ .

(2) The estimator of  $V\left(\hat{Y}_R^{\prime(3)}\right)$  is

$$\hat{V}\left(\hat{Y}_R^{\prime(3)}\right) = \sum_{i \in s} \hat{D}_i \hat{B}_i^2 + \sum_{i \in s} \sum_{i \setminus \{j\} \in s} \hat{D}_{ij} \hat{B}_i \hat{B}_j + \sum_{i \in s} \hat{E}_i \hat{B}_i^2, \quad (40)$$

where  $\hat{D}_i = \frac{r_i D_i}{\pi_i p'_i}$ ,  $\hat{D}_{ij} = \frac{r_i D_{ij}}{\pi_i p'_i}$ ,  $\hat{E}_i = \frac{r_i E_i}{\pi_i p'_i}$  and  $\hat{B}_i = x_i - \hat{R}_r \mathbf{u}'_i \hat{\boldsymbol{\beta}}_r$ ,  $p'_i = p_i$  if  $p_i$  is known

otherwise  $p'_i = \hat{p}_i$ .  $\hat{p}_i$  is the estimator of  $p_i$  from the probit or logistic regression models.

**Proof.** Let  $\hat{Y}_R^{\prime(3)}$  be defined in (38). However, the new ratio estimator  $\hat{Y}_R^{\prime(3)}$  is a function of the GREG estimator  $\hat{X}_{GREG}$  then we use the modified automated linearization approach transform  $\hat{X}_{GREG}$  to a simple form and it is defined by

$$\hat{X}_{GREG(1)} = \bar{U}' \boldsymbol{\beta} + \frac{1}{N} \sum_{i \in s} \frac{r_i C_i}{\pi_i p_i}, \quad (41)$$

where  $C_i = x_i - \mathbf{u}'_i \boldsymbol{\beta}$ . Then, the new ratio estimator  $\hat{Y}_R^{\prime(3)}$  can be approximated by,

$$\hat{Y}_{R(1)}^{\prime(3)} \cong \frac{\frac{1}{N} \sum_{i \in s} \frac{r_i y_i}{\pi_i p_i}}{\frac{1}{N} \sum_{i \in s} \frac{r_i x_i}{\pi_i p_i}} \hat{X}_{GREG(1)}. \quad (42)$$

Therefore, variance of  $\hat{Y}_R^{\prime(3)}$  can be approximated from,

$$V\left(\hat{Y}_R^{\prime(3)}\right) = V\left(N\hat{Y}_R^{\prime(3)}\right) = N^2 V\left(\hat{Y}_R^{\prime(3)}\right) \cong N^2 V\left(\hat{Y}_{R(1)}^{\prime(3)}\right). \quad (43)$$

Furthermore, the estimator of  $V\left(\hat{Y}_R^{\prime(3)}\right)$  can be obtained by,

$$\hat{V}\left(\hat{Y}_R^{\prime(3)}\right) \cong N^2 \hat{V}\left(\hat{Y}_{R(1)}^{\prime(3)}\right). \quad (44)$$

We note that  $\hat{Y}_{R(1)}^{\prime(3)}$  is a nonlinear estimator then we use steps (1) to (5) for investigating the value of  $V\left(\hat{Y}_{R(1)}^{\prime(3)}\right)$  and it is defined by,

$$V(\hat{Y}'_{R(1)}) \cong \frac{1}{N^2} \left[ \sum_{i \in U} D_i B_i^2 + \sum_{i \in U} \sum_{i \setminus \{j\} \in U} D_{ij} B_i B_j + \sum_{i \in U} E_i B_i^2 \right], \quad (45)$$

where  $B_i = x_i - \mathbf{R}u'_i \boldsymbol{\beta}$ . Substitute (45) into (43) then,

$$V(\hat{Y}'_R) \cong \sum_{i \in U} D_i B_i^2 + \sum_{i \in U} \sum_{i \setminus \{j\} \in U} D_{ij} B_i B_j + \sum_{i \in U} E_i B_i^2. \quad (46)$$

Furthermore, the estimator of  $V(\hat{Y}'_R)$  is

$$\hat{V}(\hat{Y}'_R) = \sum_{i \in s} \hat{D}_i \hat{B}_i^2 + \sum_{i \in s} \sum_{i \setminus \{j\} \in s} \hat{D}_{ij} \hat{B}_i \hat{B}_j + \sum_{i \in s} \hat{E}_i \hat{B}_i^2, \quad (47)$$

where  $\hat{D}_i = \frac{r_i D_i}{\pi_i p'_i}$ ,  $\hat{D}_{ij} = \frac{r_i D_{ij}}{\pi_i p'_i}$ ,  $\hat{E}_i = \frac{r_i E_i}{\pi_i p'_i}$  and  $\hat{B}_i = x_i - \hat{\mathbf{R}}_r u'_i \hat{\boldsymbol{\beta}}_r$ .

Next, we consider  $\hat{Y}'_R$  under the MCAR mechanism as follows. The new ratio estimator for population mean when  $\bar{X}$  is unknown and nonresponse occurs on both  $y$  and  $x$  variables under the MCAR mechanism is

$$\hat{Y}''_R = \frac{\frac{1}{N} \sum_{i \in s} r_i y_i}{\frac{1}{N} \sum_{i \in s} r_i x_i} \hat{X}'_{GREG} = \frac{\frac{1}{N} \sum_{i \in s} r_i y_i}{\frac{1}{N} \sum_{i \in s} r_i x_i} \hat{X}'_{GREG}, \quad (48)$$

where  $\hat{X}'_{GREG} = \hat{X}'_r + [\bar{U} - \hat{U}'_r]' \hat{\boldsymbol{\beta}}'_r$ ,  $\hat{X}'_r = \frac{1}{N} \sum_{i \in s} \frac{r_i x_i}{\pi_i p}$ ,  $\hat{U}'_r = \frac{1}{N} \sum_{i \in s} \frac{r_i u_i}{\pi_i p}$ ,  $\bar{U} = \frac{1}{N} \sum_{i \in U} u_i$ ,  
 $\hat{\boldsymbol{\beta}}'_r = \left( \sum_{i \in s} \frac{r_i q_i \mathbf{u}_i \mathbf{u}'_i}{\pi_i p} \right)^{-1} \left( \sum_{i \in s} \frac{r_i q_i \mathbf{u}_i w_i}{\pi_i p} \right)$ .

Then, the new ratio estimator for population mean is

$$\hat{Y}''_R = N \hat{Y}''^{(3)}. \quad (49)$$

The variance and associated estimators of  $\hat{Y}''_R$  are discussed in Lemma 5.

**Lemma 5.** Under condition (B<sub>3</sub>) with a reverse framework and where the nonresponse mechanism is MCAR.

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

$$V(\hat{Y}''_R) = \sum_{i \in U} D_i B_i^2 + \sum_{i \in U} \sum_{i \setminus \{j\} \in U} D_{ij} B_i B_j + \sum_{i \in U} E_i B_i^2, \quad (50)$$

where  $D_i = \frac{1 - \pi_i}{\pi_i}$ ,  $D_{ij} = \frac{\pi_{ij} - \pi_i \pi_j}{\pi_i \pi_j}$ ,  $E_i = \frac{1 - p}{p}$  and  $B_i = x_i - \mathbf{R}u'_i \boldsymbol{\beta}$ .

(2) The estimator of  $V(\hat{Y}''_R)$  is

$$\hat{V}(\hat{Y}''_R) = \sum_{i \in s} \hat{D}_i \hat{B}_i^2 + \sum_{i \in s} \sum_{i \setminus \{j\} \in s} \hat{D}_{ij} \hat{B}_i \hat{B}_j + \sum_{i \in s} \hat{E}_i \hat{B}_i^2, \quad (51)$$

where  $\hat{D}_i = \frac{r_i D_i}{\pi_i p'}$ ,  $\hat{D}_{ij} = \frac{r_i D_{ij}}{\pi_i p'}$ ,  $\hat{E}_i = \frac{r_i E_i}{\pi_i p'}$  and  $\hat{B}_i = x_i - \hat{\mathbf{R}}'_r u'_i \hat{\boldsymbol{\beta}}'_r$ . The value of  $p'$  is defined in (35).

**Table 1** The relative root mean square error of the new ratio estimators and associated variance estimators for  $N = 3,000$ .

Response rate (%)	$n$	The relative root mean square error of the proposed estimators				The relative root mean square error of the variance estimators			
		$\bar{X}$ is known		$\bar{X}$ is known		$\bar{X}$ is unknown		$\bar{X}$ is unknown	
		$PRMSE(\hat{Y}_R^{(2)})$		$PRMSE(\hat{Y}_R^{(3)})$		$PRMSE(\hat{V}(\hat{Y}_R^{(2)}))$		$PRMSE(\hat{V}(\hat{Y}_R^{(3)}))$	
		MAR	MCAR	MAR	MCAR	MAR	MCAR	MAR	MCAR
60	100	0.0470	0.0472	0.0465	0.0467	0.1702	0.1703	0.2763	0.3201
	200	0.0350	0.0351	0.0354	0.0361	0.1321	0.1316	0.1761	0.2069
	600	0.0322	0.0324	0.0330	0.0340	0.1150	0.1158	0.1408	0.1658
	1,200	0.0104	0.0105	0.0105	0.0107	0.0427	0.0697	0.0512	0.0804
80	100	0.0364	0.0373	0.0366	0.0375	0.1453	0.1490	0.1930	0.2526
	200	0.0258	0.0261	0.0266	0.0269	0.1127	0.1141	0.1855	0.2048
	600	0.0134	0.0139	0.0146	0.0150	0.0588	0.0614	0.1108	0.1149
	1,200	0.0086	0.0089	0.0088	0.0090	0.0433	0.0552	0.0660	0.0875

### Simulation studies

In this section, the performance of the proposed new ratio estimators and their variance estimators under the MAR mechanism is compared with the MCAR mechanism *via* simulation studies. We generated a study variable  $y_i$  from the auxiliary variables  $x_i$ ,  $w_i$ , size variable  $k_i$  and calibration variable  $u_i$  following the model from [Sichera \(2020\)](#) and it is defined by  $y_i = 0.2x_i + 0.1w_i + 2k + 3.7k^{\frac{1}{2}} + 2u_i + \varepsilon_i$  where  $k_i \sim \text{gamma}(10, 5)$ ,

$$w_i \sim \text{gamma}(5, 10), \varepsilon_i \sim N(0, 1), \begin{pmatrix} x_i \\ u_i \end{pmatrix} \sim N \left[ \begin{pmatrix} 15 \\ 5 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right], \rho = 0.70,$$

$i = 1, 2, \dots, N$ . Four levels of sample sizes  $n = 100, 200, 600$  and  $1,200$  are drawn from a population size  $N = 3,000$  and  $n = 10, 20, 60$  and  $1,200$  are drawn from a population size  $N = 300$  using [Midzuno's \(1952\)](#) scheme. We consider the MAR response mechanism with two levels of response rate; 60% and 80% and repeated the simulation 10,000 times ( $M = 10,000$ ) using Program R ([R Core Team, 2021](#)). We consider the case where the response probability is unknown and estimated by the logistic regression model for the

MAR mechanism and estimated by the function  $\hat{p} = \left( \sum_{i \in S} \frac{r_i}{\pi_i} \right) \left( \sum_{i \in S} \frac{1}{\pi_i} \right)^{-1}$  for the MCAR mechanism. The relative root mean square error (RRMSE) was used to compare the efficiency of the proposed ratio estimators and their variance estimators and the formula is

$$RRMSE(\hat{V}(\hat{Y})) = \frac{\sqrt{\frac{1}{M-1} \sum_{m=1}^M (\hat{A} - A)^2}}{A}$$

where  $\hat{A}$  is the proposed estimators or variance estimators and  $A$  is expectation of  $\hat{A}$  or  $E(\hat{A})$ . The results are shown in [Tables 1](#) and [2](#).



**Table 2** The relative root mean square error of the new ratio estimators and associated variance estimators with population size  $N = 300$ .

Response rate (%)	$n$	The relative root mean square error of the proposed estimators				The relative root mean square error of the variance estimators			
		$\bar{X}$ is known		$\bar{X}$ is unknown		$\bar{X}$ is known		$\bar{X}$ is unknown	
		$PRMSE(\hat{Y}_R^{(2)})$		$PRMSE(\hat{Y}_R^{(3)})$		$PRMSE(\hat{V}(\hat{Y}_R^{(2)}))$		$PRMSE(\hat{V}(\hat{Y}_R^{(3)}))$	
		MAR	MCAR	MAR	MCAR	MAR	MCAR	MAR	MCAR
60	10	0.1166	0.1196	0.1169	0.1198	0.6605	0.6902	2.2661	2.7669
	20	0.1038	0.1056	0.1049	0.1067	0.5423	0.5452	1.3307	1.3510
	60	0.0624	0.0626	0.0631	0.0634	0.2420	0.2489	0.3666	0.4646
	120	0.0478	0.0479	0.0483	0.0485	0.1548	0.1829	0.2500	0.2634
80	10	0.0949	0.0962	0.0957	0.0968	0.4347	0.4533	2.2106	2.5923
	20	0.0863	0.0869	0.0869	0.0871	0.3722	0.3736	1.3123	1.3319
	60	0.0480	0.0484	0.0481	0.0485	0.2084	0.2150	0.3628	0.4144
	120	0.0298	0.0301	0.0298	0.0303	0.1392	0.1514	0.2164	0.2325

The simulation results found in [Table 1](#) for  $N = 3,000$  that the new population total estimator under missing at random performed better than the estimators under missing completely at random for both situations where  $\bar{X}$  is either known or unknown. There was an increase of response rate, decrease of the relative root mean square errors as same as for the sample sizes for all estimators. When  $\bar{X}$  is unknown and needs to be estimated, it results in increasing the relative root mean square errors due to the estimation process. Similar results were found in the case of variance estimators. Similar results are found in [Table 2](#) for a smaller sample size  $N = 300$ .

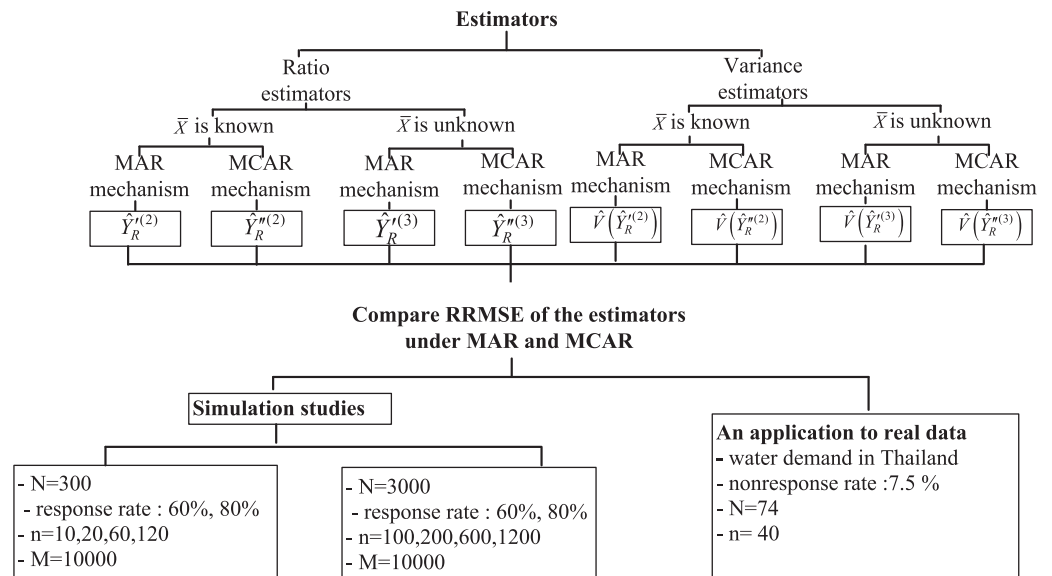
### An application to water demand in Thailand

The new estimators are applied to estimate the water demand in Thailand. The data are from the provincial waterworks during August and July 2022. *Midzuno's (1952)* scheme is instigated to select a sample of size 40 provinces from the total of 74 provinces. The demand for water in August 2022 is considered as study variable  $y$ . Two auxiliary variables  $x$  and  $w$  are the water supply in August and the water demand in July 2022, respectively. The variable  $x$  is used to construct the new ratio estimators and the variable  $w$  is used to estimate the response probabilities with the logistic regression model under the MAR mechanism. The calibration variable  $u$  is the water supply in July 2022 and the size variable  $k$  is the number of water users in August 2022. The nonresponse rate is 7.5% in this study.

[Table 3](#) shows the total estimate of water demand in August 2022, Thailand. We see that the estimated water demand when  $\bar{X}$  is known is higher than when  $\bar{X}$  is unknown under both the MAR and MCAR nonresponse mechanisms. In contrast, the estimates of variance when  $\bar{X}$  is unknown is a lot higher than the estimates of variance when  $\bar{X}$  is known due to the estimation of the unknown population mean of the auxiliary variable. The new estimators can be useful for application to the real world when nonresponse occurs in the study which requires management before the estimation process and further analysis.

**Table 3** The total estimates of water demand in August 2022.

Nonresponse mechanism	Information on the auxiliary variable	Estimated water demand	Variance estimates
MAR	$\bar{X}$ is known	122,763,533	5,621,837,076,813
	$\bar{X}$ is unknown	112,079,391	49,945,263,902,570
MCAR	$\bar{X}$ is known	122,752,240	5,958,276,721,564
	$\bar{X}$ is unknown	111,926,154	52,711,445,906,451

**Figure 1** The conclusion for all the cases of the simulation studies in an empirical study.

Full-size DOI: 10.7717/peerj.14551/fig-1

Figure 1 shows the conclusion for all the cases of the simulation studies in an empirical study.

## CONCLUSIONS

The new ratio estimators for estimating population total and population mean when missing data is missing at random occurs with both study and auxiliary variables under UPWOR when the population mean of an auxiliary variable is known and unknown are proposed. In the latter we suggested to estimate it from other variables using the GREG estimator. The new ratio estimators are compared by their efficacies under the MAR and MCAR nonresponse mechanisms through simulation studies and an empirical study using water demand data in Thailand. The results found that the new ratio estimators under the MAR mechanism are more efficient than ratio estimators under the MCAR mechanism for all response rates and sample sizes. The proposed estimators are applied to estimate the demand for water so this information can be used to plan for policies and strategies for preventing water shortages which may occur in the future. The proposed estimators are more useful in practice when compared to the estimators proposed by *Ponkaew & Lawson (2018)* that considered only under MCAR and when only the study variable is missing

which also required the known parameter of the population mean of the auxiliary variable which is difficult to find. The proposed estimators are more flexible to apply in real life because we can use them in more flexible situations when both the nonresponse mechanism is uniform or not uniform which is more likely to occur in real world problems. If the population mean of the auxiliary variable is unknown, it can be estimated using the GREG estimator which makes use of the benefit of the related variables in the estimation process to improve the efficiency of the population total estimators. We can extend the new estimator to complex survey designs such as stratified cluster sampling and consider it under the not missing at random nonresponse mechanism (NMAR).

## ACKNOWLEDGEMENTS

Thank you to all the referees for their valuable comments which help to improve the article.

## ADDITIONAL INFORMATION AND DECLARATIONS

### Funding

This research was funded by the National Science, Research and Innovation Fund (NSRF), and King Mongkut's University of Technology North Bangkok with contract no. KMUTNB-FF-65-28. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

### Grant Disclosures

The following grant information was disclosed by the authors:

National Science, Research and Innovation Fund (NSRF).

King Mongkut's University of Technology North Bangkok: KMUTNB-FF-65-28.

### Competing Interests

The authors declare that they have no competing interests.

### Author Contributions

- Chugiat Ponkaew performed the experiments, analyzed the data, prepared figures and/or tables, and approved the final draft.
- Nuanpan Lawson conceived and designed the experiments, performed the experiments, authored or reviewed drafts of the article, and approved the final draft.

### Data Availability

The following information was supplied regarding data availability:

The data and codes for the simulation studies and an application to water data are available in the [Supplemental Files](#).

### Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.14551#supplemental-information>.

## REFERENCES

- Anas MM, Huang Z, Shahzad U, Zaman T, Shahzadi S. 2022. Compromised imputation based mean estimators using robust quantile regression. *Communications in Statistics – Simulation and Computation* DOI 10.1080/03610926.2022.2108057.
- Bacanli S, Kadilar C. 2008. Ratio estimators with unequal probability designs. *Pakistan Journal of Statistics* 24(3):167–172.
- Bakker M, van Duist H, van Schagen K, Vreeburg J, Rietveld L. 2014. Improving the performance of water demand forecasting models by using weather input. *Procedia Engineering* 70:93–102 DOI 10.1016/j.proeng.2014.02.012.
- Bhushan S, Kumar A. 2022. On optimal classes of estimators under ranked set sampling. *Communications in Statistics – Theory and Methods* 51(8):2610–2639 DOI 10.1080/03610926.2020.1777431.
- Bhushan S, Kumar A, Pandey AP, Singh S. 2022. Estimation of population mean in presence of missing data under simple random sampling. *Communications in Statistics – Simulation and computation* DOI 10.1080/03610918.2021.2006713.
- Boretti A, Rosa L. 2019. Reassessing the projections of the world water development report. *npj Clean Water* 2(15):1–6 DOI 10.1038/s41545-019-0039-9.
- Chodjuntug K, Lawson N. 2022a. Imputation for estimating the population mean in the presence of nonresponse, with application to fine particle density in Bangkok. *Mathematical Population Studies* 29(4):204–225 DOI 10.1080/08898480.2021.1997466.
- Chodjuntug K, Lawson N. 2022b. The chain regression exponential type imputation method for mean estimation in the presence of missing data. *Songklanakarin Journal of Science and Technology* 44(4):1109–1118.
- Cochran WG. 1940. The estimation of the yields of cereal experiments by sampling for the ratio of grain to total produce. *The Journal of Agricultural Science* 30(2):262–275 DOI 10.1017/S0021859600048012.
- Cochran WG. 1977. *Sampling Techniques* (3rd ed.). New York: John Wiley & Sons.
- Horvitz DF, Thompson DJ. 1952. A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association* 47(260):663–685 DOI 10.1080/01621459.1952.10483446.
- Huang L, Lin L. 2017. Supply and demand analysis of water resources based on system dynamics model. *Journal of Engineering and Technological Sciences* 49(6):705–720 DOI 10.5614/j.eng.technol.sci.2017.49.6.1.
- Kaewprasert T, Niwitpong S-A, Niwitpong S. 2022. Bayesian estimation for the mean of delta—gamma distributions with application to rainfall data in Thailand. *PeerJ* 10(1):e13465 DOI 10.7717/peerj.13465.
- Khare BB, Srivastava S. 1997. Transformed ratio type estimators for the population mean in the presence of non-response. *Communications in Statistics – Theory and Methods* 26(7):1779–1791 DOI 10.1080/03610929708832012.
- Kumar S. 2015. Efficient use of auxiliary information in estimating the population ratio, product and mean in presence of non-response. *Journal of Advanced Computing* 4:68–87.
- Lawson N. 2017. Variance estimation in the presence of nonresponse under probability proportional to size sampling. In: *Proceeding of the 6th Annual International Conference on Computational Mathematics, Computational Geometry and Statistics 2017 (CMCGS 2017) 6th-7th March 2017*, Singapore.

- Lawson N. 2021.** An alternative family of combined estimators for estimating population mean in finite populations. *Lobachevskii Journal of Mathematics* **42(13)**:3150–3157 DOI [10.1134/S1995080222010115](https://doi.org/10.1134/S1995080222010115).
- Lawson N, Ponkaew C. 2019.** New generalized regression estimator in the presence of nonresponse under unequal probability sampling. *Communications in Statistics – Theory and Methods* **48(10)**:2483–2498 DOI [10.1080/03610926.2018.1465091](https://doi.org/10.1080/03610926.2018.1465091).
- Lawson N, Siripanich P. 2020.** A new generalized regression estimator and variance estimation for unequal probability sampling without replacement for missing data. *Communications in Statistics – Theory and Methods* **51(18)**:6296–6318 DOI [10.1080/03610926.2020.1860224](https://doi.org/10.1080/03610926.2020.1860224).
- Midzuno H. 1952.** On sampling system with probability proportional to sum of sizes. *Annals of the Institute of Statistical Mathematics* **3(1)**:99–107.
- Okafor FC, Lee H. 2000.** Double sampling for ratio and regression estimation with sub sampling the non-respondents. *Survey Methodology* **26**:183–188.
- Perri P. 2004.** On the efficient use of regression-in-ratio estimator in simple random sampling. In: *Atti della XLIII Riunione Scientifica della Società Italiana di Statistica*. 537–540.
- Ponkaew C. 2018.** Estimation for population total in the presence of nonresponse. PhD Dissertation, King Mongkut's University of Technology North Bangkok, Thailand.
- Ponkaew C, Lawson N. 2018.** A new ratio estimator for population total in the presence of nonresponse under unequal probability sampling without replacement. *Thai Journal of Mathematics; Special Issue (2018): Asian Conference on Fixed Point Theory and Optimization* **2018(417)**:429.
- Ponkaew C, Lawson N. 2023.** New generalized regression estimators using a ratio method and its variance estimation for unequal probability sampling without replacement in the presence of nonresponse. *Current Applied Science and Technology* **23(2)**:1–7 DOI [10.55003/cast.2022.02.23.007](https://doi.org/10.55003/cast.2022.02.23.007).
- R Core Team. 2021.** R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <https://www.R-project.org/>.
- Rao PSRS. 1986.** Ratio estimation with subsampling the nonrespondents. *Survey Methodology* **12**:217–230.
- Rao PSRS. 1987.** Ratio and regression estimates with sub-sampling the non-respondents. In: *Paper Presented at a Special Contributed Session of the International Statistical Association Meeting, Sept., 2–16. Tokyo, Japan*.
- Shahzad U, Hanif M, Sajjad I, Anas MM. 2020.** Quantile regression-ratio-type estimators for mean estimation under complete and partial auxiliary information. *Scientia Iranica* **29(3)**:1705–1715 DOI [10.24200/sci.2020.54423.3744](https://doi.org/10.24200/sci.2020.54423.3744).
- Sichera R. 2020.** Approximate inclusion probabilities for survey sampling. R package version 0.1.2. Available at <https://CRAN.R-project.org/package=jipApprox>.
- Sisodia B, Dwivedi V. 1981.** Modified ratio estimator using coefficient of variation of auxiliary variable. *Journal-Indian Society of Agricultural Statistics* **33**:13–18.
- Särndal CE. 2007.** The calibration approach in survey theory and practice. *Survey Methodology* **33(2)**:99–119.
- Särndal CE, Lundström S. 2005.** *Estimation in surveys with nonresponse*. New York: John Wiley & Sons.
- Särndal CE, Swensson B, Wretman J. 1992.** *Model assisted survey sampling*. New York: Springer-Verlag.
- Upadhyaya LN, Singh HP. 1999.** Use of transformed auxiliary variable in estimating the finite population mean. *Biometrical Journal* **41(5)**:627–636 DOI [10.1002/\(SICI\)1521-4036\(199909\)41:5<627::AID-BIMJ627>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1521-4036(199909)41:5<627::AID-BIMJ627>3.0.CO;2-W).