Bol. Soc. Paran. Mat.
©SPM – E-ISSN-2175-1188
SPM: www.spm.uem.br/bspm

(3s.) **v. 2026 (44) 13** : 1–12.
ISSN-0037-8712
doi:10.5269/bspm.80957

Statistical Inference on Population Proportion in the Presence of Auxiliary Attributes: Analysis of Radiation Data

Bassant Elkalzah, Sohaib Ahmad, M. M. Abd El-Raouf, M. A. El-Qurashi and Mohammed Elgarhy

ABSTRACT: Precise estimation of population proportions is a requirement in many disciplines, particularly when direct measurement is difficult, expensive and time consuming. In this paper, we propose a more efficient statistical inference procedure for estimating population proportion in the presence of an auxiliary attributes. The method is applied to an actual radiation dataset with the underlying primary variable considered as the proportion of individuals exposed at a certain threshold value like mean, median, first quartile and third quartile. Comparative study is performed for the proposed estimators with traditional estimators under simple random sampling. Theoretical efficiency of the proposed class of estimators are analyzed and both bias and mean squared error (MSE) expressions are obtained up to the first order of approximation, which has been verified through a comprehensive empirical study. The illustrative radiation data highlight the potential benefit to public health and environmental monitoring of incorporating auxiliary attributes, when good estimates can inform policy and safely standards. This method not only decreases the estimation error, but also provides a cost-efficient counterpart to expensive data collecting schemes. Generally, the study has shown that use of auxiliary attributes in estimation is statistically sound and practically feasible approach to improve the estimates on population proportion, especially in fields such as radiation exposure assessment where there may be inadequacy of sensitive data.

Keywords: Application on radiation data, proportion estimation, generalized estimator, efficiency comparison, minimum mean squared error.

Contents

1. Introduction

Statistical inference is an important tool for making generalizations of population parameters from sample information. One useful goal in this field is estimation of population proportion, particularly for binary outcomes representing the presence or absence of a characteristic, disease, or exposure. These kind of problems appear often in several applications, such as public health, environmental monitoring and industrial quality control and the social sciences. The validity of such estimates is even more important when the potential consequences are great, as in estimating the proportion that were exposed to hazardous levels of radiation. In many applied contexts, collecting full data on the study variable (e.g., level of radiation exposure) is impractical, expensive or unethical. However, information concerning related auxiliary attributes, e.g., age, gender, occupation, or the geographic location could be available on a finer grained level. There is a potential opportunity to leverage these auxiliary characteristics, which are

2020 *Mathematics Subject Classification:* 60E05.
Submitted December 20, 2025. Published April 28, 2026.

associated with the main study variable, to increase precision of population estimates. A covariate is usually a binary or categorical independent variable that is not of interest but has known association with the dependent variable. For example, in studies investigating radiation exposure, a person occupational status (nuclear plant worker, health sector worker or member of the general population) may be a strong predictor for potential levels of exposure to ionizing radiation. Incorporating this auxiliary information via proper estimation approaches can dramatically improve the efficiency for statistical inferences.

Several ratio type and regression type estimators have been proposed such as composite estimators which are based on the auxiliary information to minimize the variance of an estimator. These estimators, which are common in estimation of means, produce effective estimates when applied to the estimation of proportions. Indeed, authors like [1,2,3,4,5] have made a few contributions in the development of auxiliary attribute based estimators under different sampling designs such as SRSWOR, stratified sampling and two-phase design. Some more related work to population proportion published in [6,7,8,9,10,11,12,13] and [14,15] is based on auxiliary attribute under different sampling strategies.

This work is motivated by the convergence of two key ideas: the growing requirement to estimate population proportion in sensitive areas (e.g. radiation exposure monitoring, etc.) accurately and availability of auxiliary information whose potential for use has not yet been fully employed. Especially relevant is the application in radiation exposure assessment, which consists of complex environmental and occupational information, usually data limited. It is in such settings that statistical efficiency matters the most, and leveraging auxiliary features can then be a useful tool. In this study we pursue the development of better estimators for population proportion using auxiliary attribute information. It specifically seeks to develop, analyze and evaluate estimators that use available population parameters (such as the proportion of the auxiliary variable) in improving estimation accuracy where a target proportion is concerned. Closed form expressions for bias and MSE are obtained under the simple random sampling without replacement. These results are useful to comprehend some theory properties of the introduced estimators and also to compare its performance with classical ones.

The theoretical results are also confirmed using empirical data by applying them into real-life radiation exposure. In this dataset, the task is to infer fraction of people exceeding radiation level. An ancillary characteristic, e.g. working in a high-risk occupation group, is utilized to build augmented estimators. Application results show that the proposed estimators perform better in terms of smaller mean squared errors and larger relative efficiency when compared with traditional sample proportion estimator. This research not only makes a contribution to the literature on auxiliary information in survey sampling, but also has practical relevance. The concrete method is theoretically reasonable and easy to implement, which is applicable in many fields if there are auxiliary attributes. This is particularly true in public health and policy, where timely and reliable exposure or disease prevalence estimates are vital for intervention and resource allocation.

In conclusion, the current study attempts to fill an important gap in estimation of population proportions by developing and validating estimators that utilize auxiliary attribute information. The strong theoretical foundation and empirical demonstration, especially in radiation data, effectively demonstrates the practical applicability of these techniques. With the continued improvement in data availability and computational methodologies, the incorporation of auxiliary information into statistical inference will continue to be a promising option for improved estimation accuracy in various practical situations.

1.1. Key Objectives

The key objectives of the present study are as follows:

- **Construction of improved estimators of population proportion**

  To develop and propose new ratio-type and regression-type estimators for estimating the population proportion by utilizing auxiliary attribute information.

- **Effective utilization of auxiliary attributes**

  To design methodologies demonstrating how binary or categorical auxiliary characteristics (such as occupation or geographic region), which are related to the study variable (e.g., radiation exposure), can be efficiently used to improve estimation precision.

- **Establishment of theoretical properties of the proposed estimators**

  To derive expressions for the bias and mean squared error (MSE) of the proposed estimators under simple random sampling without replacement, using first-order approximation.

- **Performance comparison with existing estimators**

  To evaluate and compare the performance of the proposed estimators with traditional estimators (such as the sample proportion) using theoretical efficiency measures and empirical evidence.

- **Demonstration using real radiation exposure data**

  To illustrate the practical applicability of the proposed estimation methods through an application to a real radiation exposure dataset, where auxiliary attributes are effectively exploited.

- **Quantification of efficiency gains through auxiliary information**

  To quantify the gain in efficiency by assessing variance reduction through percentage relative efficiency when auxiliary information is incorporated.

- **Support for public health and environmental decision-making**

  To develop statistically sound tools for reliable estimation of proportions in sensitive public health contexts, such as radiation exposure, where direct measurement is challenging and informed policy decisions are essential.

2. Methodology and Notations

Let us considered that we have a population $V = (V_1, V_2, \dots, V_N)$ consist of N independent units. A sample of size n has been selected from V with the help of simple random sampling without replacement. Suppose φ_{iy} and φ_{ix} represent the study and auxiliary attributes for φ_y and φ_x. It to be considered that φ_{ij} $(j = y, x)$ if units are chosen otherwise zero.

Consider $\varphi_j = \left(\frac{a_j}{n}\right)$ and $P_j = \left(\frac{A_j}{N}\right)$ denotes the sample and population. Where $a_j = \sum_{i=1}^{n} \varphi_{ij}$, and $A_j = \sum_{i=1}^{N} \varphi_{ij}$, as $j = y, x$.

$$\rho_{\varphi_y\varphi_x} = \frac{S_{\varphi_y\varphi_x}}{S_{\varphi_y}S_{\varphi_x}}, \quad S_{\varphi_y\varphi_x} = \frac{1}{N-1}\sum_{i=1}^{N}(\varphi_{iy} - P_y)(\varphi_{ix} - P_x),$$

$$S_{\varphi_y} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\varphi_{iy} - P_y)^2}, \quad S_{\varphi_x} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\varphi_{ix} - P_x)^2}, \quad C_{\varphi_j} = \frac{S_{\varphi_j}}{P_j}, \quad j = y, x,$$

$$e_0 = \left(\frac{\varphi_{i1} - P_y}{P_y}\right), \quad e_1 = \left(\frac{\varphi_{i2} - P_x}{P_x}\right),$$

$$E(e_0^2) = \lambda C_{\phi_y}^2 = \breve{U}20, \quad E(e_1^2) = \lambda C{\phi_x}^2 = \breve{U}02, \quad E(e_0e_1) = \lambda\rho\phi_y\phi_x C_{\phi_y}C_{\phi_x} = \breve{U}_{11}, \quad \lambda = \left(\frac{1}{n} - \frac{1}{N}\right).$$

3. Existing Estimators for Population Proportion

In this section, we discussed some well-known related estimators with properties based on population proportion are given by:

1. The traditional estimator for proportion with variance is given by:

$$\widehat{P}_{trad} = \widehat{P}_y, \tag{3.1}$$

$$\mathrm{Var}(\widehat{P}trad) = P_y^2\breve{U}20 \tag{3.2}$$

2. The ratio estimator recommended by [16] with bias and mean squared error are given by:

$$\widehat{P}_R = \widehat{P}_y \left(\frac{P_x}{\widehat{P}_x} \right), \tag{3.3}$$

$$\text{Bias}(\widehat{P}R) \cong P_y \left[\breve{U}20 - \breve{U}_{11} \right],$$
$$\text{MSE}(\widehat{P}R) \cong P_y^2 \left[\breve{U}20 + \breve{U}02 - 2\breve{U}11 \right].$$

3. The product estimator suggested by [17] properties are given by:

$$\widehat{P}_P = \widehat{P}_y \left(\frac{\widehat{P}_x}{P_x} \right), \tag{3.4}$$

$$\text{Bias}(\widehat{P}P) \cong P_y \left(\breve{U}11 - \frac{1}{4}\breve{U}_{02} \right),$$
$$\text{MSE}(\widehat{P}P) \cong P_y^2 \left[\breve{U}20 + \breve{U}02 + 2\breve{U}11 \right].$$

4. The regression estimator along with variance is presented as:

$$\widehat{P}_{Reg} = \widehat{P}_y + b(P_x - \widehat{P}_x), \tag{3.5}$$

Where $b = \frac{S_{yx}}{S_y S_x}$.

$$\text{Var}(\widehat{P}Reg) = \text{MSE}(\widehat{P}Reg) \cong P_y^2 \breve{U}20 \left(1 - \rho \phi_y {\phi_x}^2\right). \tag{3.6}$$

5. The authors in [18] mentioned the following exponential type estimators along with bias and mean squared error:

$$\widehat{P}_{BR} = \widehat{P}_y \exp\left(\frac{P_x - \widehat{P}_x}{P_x + \widehat{P}_x} \right), \tag{3.7}$$

$$\widehat{P}_{BP} = \widehat{P}_y \exp\left(\frac{\widehat{P}_x - P_x}{P_x + \widehat{P}_x} \right), \tag{3.8}$$

$$\text{Bias}(\widehat{P}BR) \cong P_y \left(\frac{3}{8}\breve{U}02 - \frac{1}{2}\breve{U}_{11} \right), \tag{3.9}$$

$$\text{MSE}(\widehat{P}BR) \cong P_y^2 \left(\breve{U}20 + \frac{1}{4}\breve{U}02 - \breve{U}11 \right), \tag{3.10}$$

$$\text{Bias}(\widehat{P}BP) \cong P_y \left(\breve{U}11 - \frac{1}{4}\breve{U}_{02} \right), \tag{3.11}$$

$$\text{MSE}(\widehat{P}BP) \cong P_y^2 \left[\breve{U}20 + \frac{1}{4}\breve{U}02 + \breve{U}11 \right]. \tag{3.12}$$

6. The enhanced difference type estimators for population proportion recognized by [19,20], are given by:

$$\widehat{P}_{Diff1} = D_1 \widehat{P}_y + D_2(P_x - \widehat{P}_x), \tag{3.13}$$

$$\widehat{P}_{Diff2} = [D_3 \widehat{P}_y + D_4(P_x - \widehat{P}_x)] \left(\frac{P_x}{\widehat{P}_x} \right), \tag{3.14}$$

Where D_1, D_2, D_3 and D_4 are constants.

$$D_1 = \frac{1}{1 + \breve{U}20(1 - \rho\phi_y{\phi_x}^2)},$$

$$D_2 = \frac{P_y}{P_x}\left[\frac{\breve{U}11}{1 + \breve{U}02(1 - \rho^2_{\phi_y\phi_x})}\right],$$

$$D_3 = \frac{1 - \breve{U}20}{1 - \breve{U}20 + \breve{U}20(1 - \rho\phi_y{\phi_x}^2)},$$

$$D_4 = \frac{P_y}{P_x}\left[1 + \left(\frac{\breve{U}11}{\breve{U}02^{1/2}}\right) - 2\right].$$

The properties of $\widehat{P}Diff1$ and $\widehat{P}Diff2$ are given by:

$$\text{Bias}(\widehat{P}Diff1) = \frac{-P_y\breve{U}20(1 - \rho^2_{\phi_y\phi_x})}{1 + \breve{U}20(1 - \rho\phi_y{\phi_x}^2)}, \quad (3.15)$$

$$\text{Bias}(\widehat{P}Diff2) = P_y\left[\frac{-\breve{U}20(1 - \rho^2_{\phi_y\phi_x}) - \breve{U}20^{1/2}\breve{U}02^{1/2} + \breve{U}20^{1/2}\{\rho\phi_y\phi_x - (1 - \breve{U}20)\breve{U}11\}}{1 - \breve{U}20\rho\phi_y{\phi_x}^2}\right], \quad (3.16)$$

$$\text{MSE}(\widehat{P}Diff1) = \frac{P_y^2\breve{U}20(1 - \rho^2_{\phi_y\phi_x})}{1 + \breve{U}20(1 - \rho\phi_y{\phi_x}^2)}, \quad (3.17)$$

$$\text{MSE}(\widehat{P}Diff2) = \frac{P_y^2(1 - \breve{U}20)\breve{U}20(1 - \rho\phi_y{\phi_x}^2)}{(1 - \breve{U}20) + \breve{U}20(1 - \rho^2_{\phi_y\phi_x})}. \quad (3.18)$$

4. Proposed Class of Estimators

Accurate estimation of population proportions is a cornerstone of statistical analysis in fields such as public health, environmental monitoring, and social science research. In many practical scenarios, especially those involving sensitive or hard to measure variables like radiation exposure, obtaining complete and reliable data is both challenging and resource intensive. Traditional methods, such as using the sample proportion, are often limited in efficiency particularly when the variable of interest is rare or when the sample size is constrained due to cost or accessibility issues. However, in many such situations, auxiliary information in the form of categorical or binary attributes (e.g., occupation, location, or demographic group) is often available and correlated with the main study variable. Unfortunately, this valuable information is frequently underutilized in estimating population parameters. This gap forms the central motivation of this study. By taking motivation from [21], we need to develop more efficient and practical estimation techniques that make use of these auxiliary attributes to improve the estimation of population proportions. In the context of radiation data, for example, direct radiation measurements for all individuals may be difficult to obtain, but auxiliary data such as whether a person works in a high risk occupation is usually known. Incorporating such auxiliary attributes into the estimation process can lead to significant gains in accuracy and reliability. Therefore, this study proposes and analyzes modified ratio and regression-type estimators that leverage auxiliary attributes to enhance statistical inference. The goal is not only to reduce estimation error but also to provide tools that are applicable in real-world settings where data limitations are common. By addressing these challenges, this work aims to contribute meaningfully to both statistical methodology and its application in public health and environmental risk assessment.

$$\hat{P}G = \left[D5\hat{P}y + D6(P_x - \hat{P}_x)\right] \times \exp\left[\frac{aP_x + b}{\alpha(a\hat{P}x + b) + (1-\alpha)(aPx + b)} - 1\right] \times \left[\frac{1}{2}\left\{\frac{aP_x + b}{\alpha(a\hat{P}x + b) + (1-\alpha)(aPx + b)}\right\} + \left\{\frac{\alpha(a\hat{P}x + b) + (1-\alpha)(aPx + b)}{aP_x + b}\right\}^2\right], \tag{4.1}$$

Table 1: Some members of the recommended class of estimators

Estimators	α	a	b
$\hat{P}G1 = [D5\hat{P}y + D6(P_x - \hat{P}x)]\exp\left[\frac{Px+\rho_{\phi y\phi x}}{(\hat{P}x+\rho\phi y\phi x)} - 1\right]\left[\left(\frac{1}{2}\right)\left\{\frac{P_x+\rho_{\phi y\phi x}}{(\hat{P}x+\rho\phi y\phi x)}\right\} + \left\{\frac{(\hat{P}x+\rho\phi y\phi x)}{P_x+\rho_{\phi y\phi x}}\right\}^2\right]$	1	1	$\rho_{\phi y\phi x}$
$\hat{P}G2 = [D5\hat{P}y + D6(P_x - \hat{P}x)]\exp\left[\frac{Px+C_{\phi x}}{(\hat{P}x+C\phi x)} - 1\right]\left[\left(\frac{1}{2}\right)\left\{\frac{P_x+C_{\phi x}}{(\hat{P}x+C\phi x)}\right\} + \left\{\frac{(\hat{P}x+C\phi x)}{P_x+C_{\phi x}}\right\}^2\right]$	1	1	$C_{\phi x}$
$\hat{P}G3 = [D5\hat{P}y + D6(P_x - \hat{P}x)]\exp\left[\frac{Px+\beta_{\phi x}}{(\hat{P}x+\beta\phi x)} - 1\right]\left[\left(\frac{1}{2}\right)\left\{\frac{P_x+\beta_{\phi x}}{(\hat{P}x+\beta\phi x)}\right\} + \left\{\frac{(\hat{P}x+\beta\phi x)}{P_x+\beta_{\phi x}}\right\}^2\right]$	1	1	$\beta_{\phi x}$
$\hat{P}G4 = [D5\hat{P}y + D6(P_x - \hat{P}x)]\exp\left[\frac{\beta\phi x P_x+C_{\phi x}}{(\beta_{\phi x}\hat{P}x+C\phi x)} - 1\right]\left[\left(\frac{1}{2}\right)\left\{\frac{\beta_{\phi x}P_x+C_{\phi x}}{(\beta_{\phi x}\hat{P}x+C\phi x)}\right\} + \left\{\frac{(\beta_{\phi x}\hat{P}x+C\phi x)}{\beta_{\phi x}P_x+C_{\phi x}}\right\}^2\right]$	1	$\beta_{\phi x}$	$C_{\phi x}$
$\hat{P}G5 = [D5\hat{P}y + D6(P_x - \hat{P}x)]\exp\left[\frac{C\phi x P_x+\beta_{\phi x}}{(C_{\phi x}\hat{P}x+\beta\phi x)} - 1\right]\left[\left(\frac{1}{2}\right)\left\{\frac{C_{\phi x}P_x+\beta_{\phi x}}{(C_{\phi x}\hat{P}x+\beta\phi x)}\right\} + \left\{\frac{(C_{\phi x}\hat{P}x+\beta\phi x)}{C_{\phi x}P_x+\beta_{\phi x}}\right\}^2\right]$	1	$C_{\phi x}$	$\beta_{\phi x}$

By expressing (4.1) in error terms we have:

$$\hat{P}G = \left[D5\bar{P}y(1+e0) + D_6(P_x - P_x\{1+e_1\})\right] \times \exp\left[\frac{aP_x + b}{\alpha(aP_x\{1+e_1\}+b) + (1-\alpha)(aP_x+b)} - 1\right] \times \left[\left(\frac{1}{2}\right)\left\{\frac{aP_x + b}{\alpha(aP_x\{1+e_1\}+b) + (1-\alpha)(aP_x+b)}\right\} + \left\{\frac{\alpha(aP_x\{1+e_1\}+b) + (1-\alpha)(aP_x+b)}{aP_x+b}\right\}^2\right], \tag{4.2}$$

$$\hat{P}G = [D5P_y(1+e_0) + D_6(-P_xe_1)]\exp\left[\frac{aP_x + b}{\alpha(aP_x + aP_xe_1 + b) + (aP_x + b - a\alpha P_x - \alpha b)} - 1\right] \times \left[\left(\frac{1}{2}\right)\left\{\frac{aP_x + b}{\alpha(aP_x + aP_xe_1 + b) + (aP_x + b - a\alpha P_x - \alpha b)}\right\} + \left\{\frac{\alpha(aP_x + aP_xe_1 + b) + (1-\alpha)(aP_x + b - a\alpha P_x - \alpha b)}{aP_x + b}\right\}^2\right]. \tag{4.3}$$

Solving and simplifying (4.3), up to first order of approximation, we have:
Equation (4.4) - First order approximation

$$\hat{P}G - Py = -P_y + P_yD_5 + P_ye_0D_5 - P_y\theta e_1D_5 - P_y\theta e_0e_1D_5 + \frac{5}{2}P_y\theta^2e_1^2D_5 - P_xe_1D_6 + P_x\theta e_1^2D_6. \tag{4.4}$$

By taking expectations on both sides, we obtain bias of $\hat{P}_G$:
Equation (4.5) - Bias expression

$$\text{Bias}(\hat{P}G) = -Py + P_x\theta D_6\breve{U}02 + D5\left(P_y + \frac{5P_y\theta^2\breve{U}02}{2} - Py\theta\breve{U}_{11}\right). \tag{4.5}$$

By taking expectation and squaring (4.4), we obtain expression of mean squared error of $\hat{P}_G$, which is given by:

Equation (4.6) - MSE expression

$$\begin{aligned}\text{MSE}(\hat{P}G) =& Py^2 + P_x D_6(-2P_y\theta + P_x D_6)\breve{U}02 + PyD_5(-2P_y + \theta\{-5P_y\theta + 4P_x D_6\}\breve{U}_{02}) \\ &+ 2(P_y\theta - P_x D_6)\breve{U}02 + Py^2 D_5^2(1 + 6\theta^2\breve{U}02 - 4\theta\breve{U}11 + \breve{U}_{20}).\end{aligned} \tag{4.6}$$

To obtain the optimum values of D_5 and D_6, we have minimize ((4.6):

$$D_{5(opt)} = \frac{\breve{U}02(2 + \theta^2\breve{U}02)}{2\left[-\breve{U}11^2 + \breve{U}02(1 + 2\theta^2\breve{U}02 + \breve{U}20)\right]},$$

$$D_{6(opt)} = \frac{P_y\left[2\theta^3\breve{U}02^2 - 2\breve{U}11(-1 + \theta\breve{U}11) + \theta\breve{U}02(-2 + \theta\breve{U}11 + 2\breve{U}20)\right]}{2P_x\left[-\breve{U}11^2 + \breve{U}02(1 + 2\theta^2\breve{U}02 + \breve{U}20)\right]}.$$

Putting $D_{5(opt)}$ and $D_{6(opt)}$ in (4.6), we obtain the minimum mean squared error:

$$MSE(\hat{P}G) = \frac{P_y^2\left[4\breve{U}11^2 + \breve{U}02(9\theta^4\breve{U}02^2 - 4\theta^2\breve{U}11^2 + 4(-1 + \theta^2\breve{U}02)\breve{U}20)\right]}{4\left[\breve{U}11^2 - \breve{U}02(1 + 2\theta^2\breve{U}02 + \breve{U}_{20})\right]}. \tag{4.7}$$

5. Numerical Study

This section outlines the numerical foundation for assessing and comparing the performance of proposed estimators in estimating the population proportion of individuals exposed to radiation, utilizing a known auxiliary attribute (e.g., occupation or exposure zone). We found the mean squared error and percentage relative efficiency for the purpose of comparison of the existing and suggested class of estimators. **Population-I: [Source: [22]]**

For barX $P_1 =$ Proportion $\phi_{i1} < 1$ for $(Y \leq 94.55)$ and $\phi_{i1} > 1$ for $(Y > 94.55)$
$P_2 =$ Proportion $\phi_{i2} < 1$ for $(X \leq 113)$ and $\phi_{i2} > 1$ for $(X > 113)$
$N = 20;\ n = 5;\ P_1 = 0.4;\ P_2 = 0.5;\ \lambda = 0.15;\ \rho_{\phi_1\phi_2} = -0.6123724;\ C_{\phi_1} = 1.256562;\ C_{\phi_2} = 1.025978$

For M_y $P_1 =$ Proportion $\phi_{i1} < 1$ for $(Y \leq 83.50)$ and $\phi_{i1} > 1$ for $(Y > 83.50)$
$P_2 =$ Proportion $\phi_{i2} < 1$ for $(X \leq 113)$ and $\phi_{i2} > 1$ for $(X > 113)$
$N = 20;\ n = 5;\ P_1 = 0.5;\ P_2 = 0.5;\ \lambda = 0.15;\ \rho_{\phi_1\phi_2} = -0.4;\ C_{\phi_1} = 1.025978;\ C_{\phi_2} = 1.025978$

For Q_1 $P_1 =$ Proportion $\phi_{i1} < 1$ for $(Y \leq 64)$ and $\phi_{i1} > 1$ for $(Y > 64)$
$P_2 =$ Proportion $\phi_{i2} < 1$ for $(X \leq 99.75)$ and $\phi_{i2} > 1$ for $(X > 99.75)$
$N = 20;\ n = 5;\ P_1 = 0.7;\ P_2 = 0.75;\ \lambda = 0.15;\ \rho_{\phi_1\phi_2} = -0.3779645;\ C_{\phi_1} = 0.6716605;\ C_{\phi_2} = 0.5923489$

For Q_3 $P_1 =$ Proportion $\phi_{i1} < 1$ for $(Y \leq 122)$ and $\phi_{i1} > 1$ for $(Y > 122)$
$P_2 =$ Proportion $\phi_{i2} < 1$ for $(X \leq 121)$ and $\phi_{i2} > 1$ for $(X > 121)$
$N = 20;\ n = 5;\ P_1 = 0.25;\ P_2 = 0.25;\ \lambda = 0.15;\ \rho_{\phi_1\phi_2} = -0.06666667;\ C_{\phi_1} = 1.777047;\ C_{\phi_2} = 1.777047$

Table 2: Numerical results of MSEs of all estimators using radiation data (Population–I)

Estimator	$\bar{\mathbf{X}}$	$\mathbf{M_y}$	$\mathbf{Q_1}$	$\mathbf{Q_3}$
$\hat{P}_{\text{trad}}$	0.03789474	0.03947368	0.03315789	0.02960526
$\hat{P}_R$	0.10105260	0.11052630	0.08105263	0.06315789
$\hat{P}_P$	0.02526316	0.04736842	0.03684211	0.05526316
$\hat{P}_{\text{Reg}}$	0.02368421	0.03315789	0.02842105	0.02947368
$\hat{P}_{BR}$	0.06315789	0.06513158	0.05065789	0.03898026
$\hat{P}_{BP}$	0.02526316	0.03355263	0.02855263	0.03503289
$\hat{P}_{\text{Diff1}}$	0.02063037	0.02927509	0.02686294	0.02002861
$\hat{P}_{\text{Diff2}}$	0.01825494	0.02864613	0.02635628	0.01554519
$\hat{P}_{G1}$	0.00856219	0.01673514	0.01989103	0.00711204
$\hat{P}_{G2}$	0.00893381	0.01304743	0.01934051	0.00290603
$\hat{P}_{G3}$	0.01192053	0.01759777	0.01588523	0.00533652
$\hat{P}_{G4}$	0.01025012	0.01505748	0.01825188	0.00378157
$\hat{P}_{G5}$	0.01192053	0.01759777	0.01721981	0.00665477

Table 3: Numerical results of PREs of all estimators using radiation data (Population–I)

Estimator	$\bar{\mathbf{X}}$	$\mathbf{M_y}$	$\mathbf{Q_1}$	$\mathbf{Q_3}$
$\hat{P}_{\text{trad}}$	100.0000	100.0000	100.0000	100.0000
$\hat{P}_R$	37.5000	35.71429	40.90909	46.87500
$\hat{P}_P$	150.0000	83.33333	90.00000	53.57143
$\hat{P}_{\text{Reg}}$	160.0000	119.0476	116.6667	100.4464
$\hat{P}_{BR}$	60.0000	60.60606	65.45455	75.94937
$\hat{P}_{BP}$	150.0000	117.6471	116.1290	84.50704
$\hat{P}_{\text{Diff1}}$	183.6842	134.8371	123.4336	147.8148
$\hat{P}_{\text{Diff2}}$	207.5862	137.7976	125.8065	190.4464
$\hat{P}_{G1}$	442.5825	235.8730	166.6977	416.2697
$\hat{P}_{G2}$	424.1721	302.5399	171.4427	1018.7550
$\hat{P}_{G3}$	317.8947	224.3108	208.7341	554.7671
$\hat{P}_{G4}$	369.7004	262.1533	181.6684	782.8836
$\hat{P}_{G5}$	317.8947	224.3108	192.5566	444.8732

Population-II: [Source: [23,24]]

For barX $P_1 =$ Proportion $\phi_{i1} < 1$ for $(Y \leq 106)$ and $\phi_{i1} > 1$ for $(Y > 106)$
$P_2 =$ Proportion $\phi_{i2} < 1$ for $(X \leq 106.1)$ and $\phi_{i2} > 1$ for $(X > 106.1)$
$N = 20;\ n = 5;\ P_1 = 0.4;\ P_2 = 0.5;\ \lambda = 0.152381;\ \rho_{\phi_1\phi_2} = -0.6123724;\ C_{\phi_1} = 1.256562;\ C_{\phi_2} = 1.025978$

For M_y $P_1 =$ Proportion $\phi_{i1} < 1$ for $(Y \leq 99)$ and $\phi_{i1} > 1$ for $(Y > 99)$
$P_2 =$ Proportion $\phi_{i2} < 1$ for $(X \leq 91)$ and $\phi_{i2} > 1$ for $(X > 91)$
$N = 21;\ n = 6;\ P_1 = 0.4761905;\ P_2 = 0.4285714;\ \lambda = 0.1190476;\ \rho_{\phi_1\phi_2} = -0.8257228;\ C_{\phi_1} = 1.074709;\ C_{\phi_2} = 1.183216$

For Q_1 $P_1 =$ Proportion $\phi_{i1} < 1$ for $(Y \leq 95)$ and $\phi_{i1} > 1$ for $(Y > 95)$
$P_2 =$ Proportion $\phi_{i2} < 1$ for $(X \leq 73)$ and $\phi_{i2} > 1$ for $(X > 73)$
$N = 21;\ n = 6;\ P_1 = 0.6666667;\ P_2 = 0.7142857;\ \lambda = 0.1190476;\ \rho_{\phi_1\phi_2} = -0.4472136;\ C_{\phi_1} = 0.7245688;\ C_{\phi_2} = 0.6480741$

For Q_3 P_1 = Proportion $\phi_{i1} < 1$ for $(Y \leq 121)$ and $\phi_{i1} > 1$ for $(Y > 121)$
P_2 = Proportion $\phi_{i2} < 1$ for $(X \leq 124)$ and $\phi_{i2} > 1$ for $(X > 124)$
$N = 21;\ n = 6;\ P_1 = 0.2380952;\ P_2 = 0.1428571;\ \lambda = 0.1190476;\ \rho_{\phi_1\phi_2} = -0.2282177;\ C_{\phi_1} = 1.83303;\ C_{\phi_2} = 2.50998$

Table 4: Numerical results of MSEs of all estimators using radiation data (Population–II)

Estimator	$\bar{\mathbf{X}}$	$\mathbf{M_y}$	$\mathbf{Q_1}$	$\mathbf{Q_3}$
$\hat{P}_{\text{trad}}$	0.02947846	0.03117914	0.03555556	0.02902494
$\hat{P}_R$	0.11111110	0.12566140	0.09244444	0.05079365
$\hat{P}_P$	0.03854875	0.01228269	0.03555556	0.01451247
$\hat{P}_{\text{Reg}}$	0.02222222	0.00992064	0.02844444	0.00634921
$\hat{P}_{BR}$	0.05895692	0.06897203	0.05688889	0.03900227
$\hat{P}_{BP}$	0.02267574	0.01228269	0.02844444	0.02086168
$\hat{P}_{\text{Diff1}}$	0.01927130	0.00950480	0.02673350	0.00570972
$\hat{P}_{\text{Diff2}}$	0.02106568	0.00975568	0.02616889	0.00295650
$\hat{P}_{G1}$	0.00134629	0.00365090	0.01995079	0.00319063
$\hat{P}_{G2}$	0.00342438	0.00365090	0.01851842	0.00341630
$\hat{P}_{G3}$	0.00758083	0.00706501	0.01671009	0.00313400
$\hat{P}_{G4}$	0.00288346	0.00967284	0.01835186	0.00338995
$\hat{P}_{G5}$	0.00919856	0.00719440	0.01782113	0.00326579

Table 5: Numerical results of PREs of all estimators using radiation data (Population–II)

Estimator	$\bar{\mathbf{X}}$	$\mathbf{M_y}$	$\mathbf{Q_1}$	$\mathbf{Q_3}$
$\hat{P}_{\text{trad}}$	100.0000	100.0000	100.0000	100.0000
$\hat{P}_R$	26.53061	24.81203	38.46154	57.14133
$\hat{P}_P$	76.47059	253.8462	100.0000	200.0000
$\hat{P}_{\text{Reg}}$	132.6531	314.2857	125.0000	457.1429
$\hat{P}_{BR}$	50.00000	45.20548	62.50000	74.42069
$\hat{P}_{BP}$	130.0000	253.8462	125.0000	139.1283
$\hat{P}_{\text{Diff1}}$	152.9656	328.0357	133.0000	508.2490
$\hat{P}_{\text{Diff2}}$	139.9360	319.5997	135.8696	981.6495
$\hat{P}_{G1}$	2189.601	854.0135	178.2163	909.7160
$\hat{P}_{G2}$	860.8411	854.0135	192.0010	849.5345
$\hat{P}_{G3}$	388.8554	441.3174	212.7790	926.2295
$\hat{P}_{G4}$	1022.328	322.3371	193.7436	856.2558
$\hat{P}_{G5}$	320.4682	433.3805	199.5134	889.2331

6. Discussion

As mentioned above, Table 1, include some members of the recommended class of estimators. The result of mean squared error and percentage relative efficiency based on population-I are presented in Table 2 and Table 3. Similarly result of all considered estimators for the mean squared error and percentage relative efficiency based on population-II are given in Tables 4 and 5. The sensitivity, expense and technical factors involved in direct measurements make them unsuitable for the estimated of population proportions regarding radiation exposure. The actual percentage of potentially exposed individuals is difficult to discern from radiation exposure data that may be subject to incomplete reporting, small sample sizes or biased sampling. The results of this study have important implications with regard to the necessity and usefulness of using auxiliary attribute information in improving the accuracy of such estimates. The use of supplementary features like occupation group, residence near radiation source and frequency of radiation-related task are using additional information which is associated with the exposure in question. It is shown that the estimators based on adding these auxiliary variables, uniformly dominate the standard sample proportion estimator in the sense of bias reduction and mean squared error. This efficiency increase is most pronounced when the radiation dataset consists of rare or cluster-localized direct exposures. Because these ratio-type and regression-type estimators account for the known population proportion of the auxiliary characteristic, they can decrease variance associated with a simple random sample drawn from binary exposure categorizations. For example, in radiation workers, occupation correlates closely with exposure so that by including the known occupational proportion, the estimator can "borrow strength" from this relationship to improve upon inference over all exposed people. Practically, these enhancements are worth their weight in gold. The correct estimate of the prevalence of exposures is important for public health policymakers so that they can focus surveillance efforts, provide resources to healthcare providers and enact protective measures. The proposed approach is cost-effective as compared to the impractical direct measurement campaigns that may be expensive and/or logistically prohibitive. In addition, the empirical results show that the improvement in efficiency is very sensitive to the strength of association between study variable and auxiliary attribute. Where this association is weak, the benefit of auxiliary information is reduced, underscoring the need for judicious choice and validation of ancillary data in study planning. In addition to contributing to the advancement of statistical methodology for estimating population proportion, this work provides applied solutions for radiation science research.

7. Conclusion

In this paper, two new ratio-type and regression-type estimators are introduced for precise estimation of the population proportion through efficient use of auxiliary attribute information. Theoretical derivations reveal that these estimators have lower bias and mean squared error than the ordinary sample proportion estimator in simple random sampling without replacement. The practical value of the method is demonstrated on radiation exposure measurements. Making use of easily collected secondary covariates, e.g., job categories associated with radiation exposure, the estimators achieve a significant gain in precision for estimating the proportion of exposed individuals. This improvement is especially important in radiation experiments, for which the direct measurement can be expensive, limiting or challenging. The empirical findings reveal that the extent of efficiency gain is significantly related to the correlation between the auxiliary variable and exposure variable. In the presence of a strong relationship, the new estimators represent significant advancements leading to more accurate population inference. This in turn, enable more informed public health decisions, risk estimations and resource distribution in radiation protection. It emphasizes the importance of ancillary data for strengthening the statistical inference and is an easily applicable, cost-effective mechanism in environmental/occupational health research. These methods could be extended in future work to more complex sampling designs and to other settings with auxiliary information, thereby further increasing the range of where this approach can be applied.

Future Recommendations

- **Extension to complex sampling designs:** The proposed general estimators can be extended to more complex and realistic sampling designs such as stratified, cluster, or systematic sampling, which are commonly employed in large-scale radiation exposure surveys. Such extensions would enhance the applicability of the estimators in practical survey settings.
- **Incorporation of multiple auxiliary attributes:** Considering the joint utilization of multiple auxiliary attributes or incorporating continuous auxiliary variables may further improve estimation accuracy. Future work may focus on the development of multivariate ratio or regression-type estimators under mixed-type auxiliary information.
- **Robustness to measurement errors:** Radiation exposure data and associated covariates may be subject to measurement errors or misclassification. The development of robust and high-dimensional estimation techniques capable of handling such misspecifications would lead to more reliable and refined statistical inference.
- **Adaptation for small area estimation:** Adapting the proposed methodology for small area estimation can facilitate reliable inference for small geographic or demographic subpopulations where direct sample information is limited or unavailable, thereby supporting targeted risk assessment and intervention strategies.
- **Integration into spatial and temporal models:** Incorporating spatial and temporal correlations in radiation exposure through the integration of auxiliary information with spatial–temporal statistical models may improve dynamic monitoring, prediction, and risk evaluation.

- **Software development for practical implementation:** The development of user-friendly software packages implementing the proposed estimators would enhance accessibility and usability for public health researchers, practitioners, and policymakers.
- **Validation on diverse real-world datasets:** Further empirical validation using diverse real-world radiation exposure datasets, including medical, industrial, and environmental contexts, would strengthen the generalizability and practical utility of the proposed estimation approaches.

Conflict of Interest

The authors declare that there is no conflict of interest.

Data Availability

All data used in this study are available within the manuscript.

References

1. Alghamdi, S. M., Ahmad, S., Almarzouki, S. M., Aloraini, B., Badr, M. M., and Abdelkawy, M. A., *Constructing a new estimator for estimating population mean utilizing auxiliary information in probability proportional to size sampling*, Alexandria Engineering Journal, 110, 506–511, (2025). DOI: https://doi.org/10.1016/j.aej.2024.12.023
2. Abd-Elfattah, A. M., El-Sherpieny, E. A., Mohamed, S. M., and Abdou, O. F., *Improvement in estimating the population mean in simple random sampling using information on auxiliary attribute*, Applied Mathematics and Computation, 215, 12, 4198–4202, (2010). DOI: https://doi.org/10.1016/j.amc.2009.12.009
3. Hussain, S., Akhtar, S., and El-Morshedy, M., *Modified estimators of finite population distribution function based on dual use of auxiliary information under stratified random sampling*, Science Progress, 105(3), 00368504221128486, (2022). DOI: https://doi.org/10.1177/00368504221128486.
4. Bhushan, S., and Kumar, A., *Evaluating the performance of logarithmic type estimators using auxiliary attribute*, Life Cycle Reliability and Safety Engineering, 12, 4, 285–292, (2023). DOI: https://doi.org/10.1007/s41872-023-00237-8
5. Ahmad, S., and Shabbir, J., *Proportion estimation using enhanced class of estimators under simple random sampling: Application with real data sets and simulation*, Measurement: Interdisciplinary Research and Perspectives, (2024). DOI: https://doi.org/10.1080/15366367.2024.2324119
6. Hussain, S., Ahmad, S., Akhtar, S., Javed, A., and Yasmeen, U., *Estimation of finite population distribution function with dual use of auxiliary information under nonresponse*, PLoS ONE, 15(12), e0243584, (2020). DOI: https://doi.org/10.1371/journal.pone.0243584
7. Saini, M., Jitendrakumar, B. R., and Kumar, A., *Optimum estimator in simple random sampling using two auxiliary attributes with application in agriculture, fisheries and education sectors*, MethodsX, 9, 101915, (2022). DOI: https://doi.org/10.1016/j.mex.2022.101915
8. Semary, H. E., Ahmad, S., Elbatal, I., Chesneau, C., Elgarhy, M., and Almetwally, E. M., *New comprehensive class of estimators for population proportion using auxiliary attribute: Simulation and an application*, Alexandria Engineering Journal, 99, 130–136, (2024). DOI: https://doi.org/10.1016/j.aej.2023.10.041
9. Singh, H. P., and Solanki, R. S., *Improved estimation of population mean in simple random sampling using information on auxiliary attribute*, Applied Mathematics and Computation, 218, 15, 7798–7812, (2012). DOI: https://doi.org/10.1016/j.amc.2011.12.053
10. Hussain, S., Ahmad, S., Saleem, M., and Akhtar, S., *Finite population distribution function estimation with dual use of auxiliary information under simple and stratified random sampling*, PLoS ONE, 15(9), e0239098, (2020). DOI: https://doi.org/10.1371/journal.pone.0239098
11. Zamanzade, E., and Wang, X., *Estimation of population proportion for judgment post-stratification*, Computational Statistics & Data Analysis, 112, 257–269, (2017). DOI: https://doi.org/10.1016/j.csda.2017.03.006
12. Hussain, S., Zichuan, M., Hussain, S., Iftikhar, A., Asif, M., Akhtar, S., and Ahmad, S., *On estimation of distribution function using dual auxiliary information under nonresponse using simple random sampling*, Journal of Probability and Statistics, 2020, Article ID 1693612, (2020). DOI: https://doi.org/10.1155/2020/1693612
13. Zaman, T., and Kadilar, C., *Novel family of exponential estimators using information of auxiliary attribute*, Journal of Statistics and Management Systems, 22, 8, 1499–1509, (2019). DOI: https://doi.org/10.1080/09720510.2019.1668195
14. Zaman, T., *New family of estimators using two auxiliary attributes*, International Journal of Advanced Research in Engineering & Management (IJAREM), 4(11), 11–16, (2018).
15. Zaman, T., and Kadilar, C., *Novel family of exponential estimators using information of auxiliary attribute*, Journal of Statistics and Management Systems, 22(8), 1499–1509, (2019).
16. Cochran, W. G., *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, (1940). DOI: https://doi.org/10.1017/S0021859600048012
17. Murthy, M. N., *Product method of estimation*, Sankhyā: The Indian Journal of Statistics, Series A, 26, 69–74, (1964).
18. Bahl, S., and Tuteja, R., *Ratio and product type exponential estimators*, Journal of Information and Optimization Sciences, 12, 1, 159–164, (1991). DOI: https://doi.org/10.1080/02522667.1991.10698946
19. Rao, T. J., *On certain methods of improving ratio and regression estimators*, Communications in Statistics–Theory and Methods, 20, 10, 3325–3340, (1991). DOI: https://doi.org/10.1080/03610929108830664
20. Gupta, S., Shabbir, J., and Ahmad, S., *Estimation of median in two-phase sampling using two auxiliary variables*, Communications in Statistics–Theory and Methods, 37, 11, 1815–1822, (2008). DOI: https://doi.org/10.1080/03610920701712105

21. Manshi, T., Kayid, M., and Alshehri, M. A., *A generalized class of estimators for population proportion using auxiliary attributes: Applications to radiation data sets*, Journal of Radiation Research and Applied Sciences, 18, 4, 102003, (2025). DOI: https://doi.org/10.1016/j.jrras.2025.102003

22. El-Saeed, A. R., Ahmad, S., and Aloraini, B., *Robust estimator for estimation of population mean under PPS sampling: Application to radiation data*, Journal of Radiation Research and Applied Sciences, 18(2), 101384, (2025).

23. Aloraini, B., Diab, L. S., and Ahmad, S., *Regression cum exponential type estimator for population proportion using auxiliary attributes: Application on radiation data and simulation study*, Journal of Radiation Research and Applied Sciences, 18(2), 101479, (2025).

24. Alomani, G., Kayid, M., and Ahmad, S., *Application to radiation data sets by suggesting an improved mean estimator under probability proportional to size sampling*, Journal of Radiation Research and Applied Sciences, 18(2), 101354, (2025).

Bassant Elkalzah,
Department of Mathematics and Statistics,
College of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU),
Riyadh, 11432, Saudi Arabia;
Department of Statistics, Mathematics and Insurance, Faculty of Business, Alexandria University,
Alexandria, 21526, Egypt.
E-mail address: belkalzah@imamu.edu.sa
E-mail address: bassant.elkalzah@alexu.edu.eg

and

Sohaib Ahmad,
Department of Statistics,
Abdul Wali Khan University,
Mardan, Pakistan.
E-mail address: sohaib_ahmad@awkum.edu.pk

and

M. M. Abd El-Raouf,
Institute of Basic and Applied Science,
College of Engineering and Technology,
Arab Academy for Science, Technology and Maritime Transport,
P.O. Box 1029, Abu Quir Campus, Alexandria, Egypt
E-mail address: m_abdelraouf85@aast.edu

and

M. A. El-Qurashi,
Institute of Basic and Applied Science,
College of Engineering and Technology,
Arab Academy for Science, Technology and Maritime Transport,
P.O. Box 1029, Abu Quir Campus, Alexandria, Egypt
E-mail address: m.elqurashi@aast.edu

and

Mohammed Elgarhy,
Faculty of Computers and Information Systems,
Egyptian Chinese University,
Nasr City, Egypt;
Department of Basic Sciences,
Higher Institute of Administrative Sciences,
Belbeis, AlSharkia, Egypt;
Department of Computer Engineering,
Biruni University, 34010, Istanbul, Turkey.
E-mail address: dr.moelgarhy@gmail.com