Int. J. Anal. Appl. (2026), 24:192

International Journal of Analysis and Applications

Development of an Improved Comprehensive Estimator for Population Proportion: Evidence from Radiation Data and Simulation Studies

Bassant Elkalzah[1,2,*], Yasser Salah S. Abougamea[3], Abdelfatah Abasher[3], Sohaib Ahmad[4], Mohammed Elgarhy[5,6]

[1]*Department of Mathematics and Statistics, College of Science, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh 11432, Saudi Arabia*

[2]*Department of Statistics, Mathematics and Insurance, Faculty of Business, Alexandria University, Alexandria 21526, Egypt*

[3]*Department of Mathematics, College of Science, Jazan University, P.O. Box 114, Jazan 45142, Kingdom of Saudi Arabia*

[4]*Department of Statistics, Abdul Wali Khan University, Mardan, Pakistan*

[5]*Faculty of Computers and Information Systems, Egyptian Chinese University, Nasr City, Egypt*

[6]*Department of Computer Engineering, Biruni University, 34010 Istanbul, Turkey*

**Corresponding author:* bassant.elkalzah@alexu.edu.eg

Abstract. Accurate population proportion estimation is a fundamental requirement in survey sampling, practicality when direct measurement is limited by cost, time or other reasons. Use of auxiliary variable has been proposed as a feasible strategy for improving estimation efficiency. The purpose of this paper is to suggest an improved class of generalized estimators for population proportion using auxiliary attribute under simple random sampling without replacement. The proposed class of estimators is analytically derived under the minimum mean squared error criterion, and its theoretical properties such as bias and MSE are obtained up to the first order of approximation. To be specific, this general form is able to include such known estimators as special cases and therefore provides a unified and flexible estimation procedure. Comparative theoretical studies are established under which the new class has improved relative efficiency as compared to the usual estimators. To illustrate the empirical value of the proposed method, we apply it to actual radiation data and a simulation study is used. The empirical results indicate that the suggested estimators provide significant improvement in efficiency, especially when the auxiliary attribute are correlated with the study attribute. These findings emphasize that the estimator is robust and applicable to a variety of practical settings. The work highlights for the general statistical theory that estimation of population proportion may be improved by more effective use of auxiliary attributes.

Received: Feb. 28, 2026.

2020 *Mathematics Subject Classification.* 62D05.

Key words and phrases. radiation data science; simulation study; generalized estimator; study and auxiliary attributes; efficiency.

https://doi.org/10.28924/2291-8639-24-2026-192
ISSN: 2291-8639

1. Introduction

Population proportion estimation is a fundamental problem in survey sampling, in particular for fields where binary states such as presence/absence, exposed/not-exposed and compliant/non-compliant are focused on. In several applied situations, including epidemiology, environmental monitoring studies, social sciences and radiologic testing, the quantity of interest is quantitatively not a continuous response but a proportion which gives the prevalence of some characteristic in the population under study. However, estimating such proportions accurately can be methodologically and operationally difficult, if not impossible when direct observation is costly, dangerous or infeasible. In these settings, auxiliary information more specifically auxiliary binary or categorical data that are associated with the target variable of interest provides a powerful tool through which estimation efficiency can be improved. Secondary properties are additional features that can be more easily observed or that are population level characteristics and for which a statistical association with the study variable is present. When properly utilized, such information provides for significantly decrease in the variance or MSE of estimators, and therefore increases the accuracy and credibility when drawing inferences from finite samples.

The inclusion of auxiliary information into estimators is a tradition in statistical sampling theory. Based on original works, ratio, product, regression and exponential-type estimators which use auxiliary information to enhance the estimations of population means or totals have been suggested. Recently, general class of estimators covering many existing methods as special cases have also been introduced and they can flexibly cover various data structures and sampling scenarios. Indeed, authors like [3], [7], [8], [9], [10], and [11] 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. Not with standing these methodological improvements, several problems persist. First, many of the existing estimator classes are based on strong assumptions about the information that is available (and accurate) with respect to auxiliary population parameters like the proportion of units whose have the auxiliary feature. In reality, this kind of information may not be easily accessible, obsolete or subject to inaccuracies. Second, much of the previous work mainly deal with theoretical development and simulation-based evaluation, not being applied to real-world data sets involved in environmental or public health monitoring, which have various challenges. Third, and most importantly the use of such estimators has been relatively little considered in radiation studies, where estimation issues are particularly serious because direct data measurements may carry possible risks.

Motivated by these requirements, this paper constructs a new general class of estimators for population proportion making use of auxiliary information. Our contributions are threefold. We first develop a generalized estimating approach under simple random sampling without replacement that incorporates auxiliary attributes in flexible manner. Second, we derive expressions for the bias and MSE of the proposed class to the first-order, and analytically characterize optimal members within the class (i.e. by finding parameter values that minimize MSE). Third, we present an empirical performance of our estimators on actual radiation data and simulation analysis compared with other known estimators. The empirical analysis sheds light not only on relative efficiency gains but also practical matters such as how strong the auxiliary covariate must be, how large sample sizes need to be and what should we expect about estimator performance in the presence of real data features (e.g., skewness, measurement error, heterogeneity).

The study provides a short review on existing method of estimation for different kinds of auxiliary variables is depicted their theoretical basis and usefulness. In Section 3 we also introduce the generalized estimator class, describing in detail its mathematical expression, assumptions and derivation of bias and MSE. In Section 4, we investigate the choice of hyper parameter values within the estimator class and show that under some conditions theoretical superiority can be reached with respect to standard estimators. We present in Section 5 empirical applications with radiation data sets, showing the relative efficiency and estimation accuracy of the proposed estimators compared to existing approaches. Finally, Section 6 concludes with discussion of practical implications and directions for future research. This paper is about improving the literature in proportion estimation informed by motivating theory and validated empirically while practical for use that leverages auxiliary attribute information in reasonable and effective manners. The proposed methodology is of great value not only for radiation control but also in other environmental sciences, epidemiology and public health.

Many estimators claim that researchers have used more information to find out the population proportions over the years. The classical ratio estimator, for instance, is the ratio of the means of a quantitative auxiliary variable from the population and the sample. Where auxiliary information is qualitative (binary or categorical) authors have further extended these concepts to construct ratio-type and regression-type estimators based on attributes. A well-known effort on this has been to derive formulas that exploit the known proportion of an alternate variable in the population to better the proportion of the variable of study. These estimators have frequently produced less biased and lower MSE estimators under certain circumstances, e.g. when there is a strong correlation between the auxiliary and the study variable. Some published work in [12–19] and [23, 24] is based on auxiliary attribute under different sampling strategies.

1.1. **Application to radiation data.**

Radiation data science involves an interdisciplinary framework for the collection, analysis and interpretation of information about ionizing radiation exposure, radiological surveillance, environmental contamination and public health risk assessment. The assessment of population-level proportion, for example, the proportion of people or places exceeding some level of radiation and materials requires estimation and is part of the science underlying radiological monitoring, protection and decision-support. However, it is difficult to obtain the actual radiation data directly because of logistical and economic reasons in non-emergency situations or safety considerations after accidents as well as the scope extent when dealing with extensively affected regions of examination on a map.

In such situations, utilization of auxiliary attribute information in estimation such as is the case with synthetic data offers an attractive alternative to statistical inference that is both more precise and accurate. Ancillary data like land use types, and proximities to atomic radiation piles or protecting facilities, or historical exposure information may be obtained via administrative databases as well as remote sensors. Although these are not direct measurements of radiation exposure, their relationship with the study variable may be used to improve estimator efficiency.

The enhanced generalized class of estimators considered in this paper is particularly convenient for such scenarios, as well. Constructed in the design-based simple random sample without replacement framework, the estimator can efficiently utilize any auxiliary categorical subject information to adjust for bias and minimize both mean-square error when estimating population proportions. To illustrate the relevance of this approach, we use it to a real exposure dataset from an area with spatial heterogeneity on contamination risk. The binary study variable indicates the category of the observation unit that has, or not exceed a given radiation safety standard, and the auxiliary variables are geographic zoning (fragmentation), e.g., residential/industrial/agricultural land use policy area; proximity to known contamination site.

It is empirically demonstrated that the general proposed estimator dominates over its ordinary competitors, among others (such as sample proportion and ratio-type estimators) in terms of several criteria. It is interesting to observe here that high relative efficiency can also be observed for moderate and strong correlation between the auxiliary attributes and the target variable. For instance, "industrial" land uses and distances from contaminated facilities were over-represented in sites with excess radiation. The inclusion of such covariates within the estimation process allows the proposed estimator to provide greater precision without requiring an increase in sample size an advantage of considerable applied interest for resource-limited field studies.

In order to test the robustness of the results, sensitivity analyses were performed by changing the magnitude of association between the auxiliary variables and study variable. The findings verify that in the case of less correlation the estimator offers better properties than its competitors. The generic structure also allows optimization of parameters, which can adjust the estimator to any particular data property of a radiation monitoring campaign: e.g., attribute misclassification, spatial clustering and measurement errors.

These results highlight the methodological and practical importance of the proposed estimator in radiation data science. So in particular, it brings really significant improvements in several aspects:

(1) Estimates of the proportion of people who might be exposed to harmful levels of radiation would allow for specific health interventions, medical monitoring, and long term epidemiological monitoring.
(2) Determining the prevalence of contamination across ecological types helps and land detailing prioritization for land reclamation planning and policy development.

(3) Statistically efficient estimating procedures can be used by regulatory agencies to assess compliance with exposure limits based on smaller or more strategically collected samples, thus costs are minimized while maintaining reliability.
(4) The rapid estimation of affected populations using such alternative spatial or demographic data in radiological emergencies can be utilized to guide evacuation and resource allocation.

Overall, our estimator transcends more than a mere theoretical novelty by providing us with a universal, flexible instrument for practical uses in radiation science. Through incorporating auxiliary attribute data (which is often ignored in practice) into the core estimation pipeline, it can help to strengthen both the statistical credibility and operational viability of radiological assessment campaigns. The need for this integration can be particularly important in areas where the consequences of errors in estimates are extensive and measurement cost is high.
Therefore, aside from the importance of issue addressed in this study to advance sampling theory itself, application suggestions for radiation scientists, environmental statisticians and public health authorities are provided. The proposed framework introduces a more objective evidence based and emphasizes the importance of modern statistical methodologies in radiation data science.

1.2. **Research Gap.** Despite substantial progress in survey sampling, especially with respect to use of auxiliary information for estimating population characteristics, there still remain a few significant challenges that hinder the extension and adoption of useful estimators for population proportions particularly applied within radiation data science.
First, the existing literature is highly focused on estimation of population means and totals related to continuous study variables and has paid little attention to estimation of population proportions, an important parameter in public health, environmental monitoring, and radiological risk assessment. For the few available methods, there are usually some limitations (e.g., based on ordinary ratio-type or linear type estimators) that may not take full advantage of auxiliary attribute information, especially when such attributes are categorical or binary.
Second, although certain classes of the general structure-based estimators with auxiliary information are studied, they often have a non flexible structural assumption or work only for specific types of data environments. Oftentimes such estimators fail to be flexible enough for complex relationships among the study and auxiliary variables or do not easily allow optimization for minimum bias and mean squared error under realistic sampling conditions.
Even more crucial is the fact that lack of experimental validation over real world data for these estimator. A considerable amount of the previous literature heavily depends on synthetic or simulated data in order to illustrate the theoretical gains in efficiency when, given a specific design, spatial heterogeneity of measurement accuracy and between-place conditional variability (contextual effects) are neglected systematically over applied settings—of which radiation-related data is representative. Within radiation science, where direct measurement of exposure can be hampered by considerations of safety, expense and logistics, the potential utility of such estimators has yet to be realized.

Moreover, this is seldom accounted for in conventional methods which do not account for misclassification or uncertainty in auxiliary attributes known to be prevalent in field-based radiation studies. For instance, indirect measures of exposure such as land-use type, distance from radiation sources or historical exposure zones could be only imperfectly associated with the true level of exposure and this would add a further layer of complexity in prediction that current methods are not able to deal with.

In light of these limitations, a natural and urgent question arises to develop a class of estimators for population proportion that is: (i) statistically sound, (ii) computationally lean, (iii) general purpose, and (iv) empirically secured:

This paper intends to fill in these important gaps by suggesting a more generalized estimator class specific to population proportion estimation under simple random sampling without replacement, establishing its statistical properties under first order approximation and further checking the finite sample performance through applying empirical radiation exposure data. In the process, it helps enrich not only the theoretical investigation of sampling estimators but also their application, particularly in fields where accurate estimation is required and data acquisition is inevitably restricted.

1.3. **Novelty of the Work.** In this article, we contribute to survey sampling theory by proposing a new generalized class of estimators for estimating population proportion under the simple random sampling without replacement when auxiliary information fails to exist. In contrast to classical estimators that are largely based on continuous auxiliary variables or may be subject to restrictive assumptions, the present approach is novel and makes clear use of binary and categorical types of auxiliary attributes, which leads to a substantial increase in the flexibility and practical relevance of the use of auxiliary information for proportion estimation.

One major contribution is to introduce a joint framework for estimator design which generalizes several well-known estimators and analytically derive explicit expressions for bias and mean squared error based on first-order approximation. Such analytical tractability allows us to systematically optimize the estimator parameters for convergence to the lowest possible estimation error, a dimension ignored in much of existing literature on proportion estimation.

Moreover, the current article also significantly contributes to the literature by connecting recently improved theories with empirical application on real radiation exposure data sets using the developed estimator. This domain in applications is a rigorous test bed with inherent measurement constraints and obscure auxiliary attribute patterns. Estimators are empirically evaluated across an extensive range of scenarios to show robustness, efficiency and adaptability of the proposed one compared with classical approaches; thus indicating its crucial importance for a broad set of applications in fields ranging from radiological risk assessment, environmental surveillance and public health monitoring.

In conclusion, what makes this research original is its complete interplay between elegance of theory and methodology with scrupulous empirical verification achieving a new standard for the accurate and efficient estimation of population proportions using auxiliary covariates under complex applied circumstances where resources are limited.

1.4. **Objectives of the Study.** The principal objectives of this study are delineated as follows:
To develop an improved general class of estimators in estimating the population proportion by ingeniously incorporating auxiliary attribute information such as categorical and binary characteristics into the structure of simple random sample without replacement.
To demonstrate, under rigorous mathematical derivation, the statistical properties of our proposed estimator such as explicit expressions for bias and MSE at first-order approximation that is critical for complete theoretical analysis.
To detect and characterize the optimal configuration of parameters that reduce error in estimation, leading to ever-increasing efficiency and precision of the estimator.
To confirm empirically that the proposed estimator is effective and robust by applying it real radiation exposure datasets with complex auxiliary structures and practical measurement constraints.
To compare the proposed estimator with existing classical estimates on the basis of accuracy and performance, and to illustrate its improved robustness in radiological risk assessment and environmental surveillance.
To develop a flexible and scalable estimation platform that is easily extensible or specialized to different applied research domains with known auxiliary attribute data, in order to connect methodology development to practical use.

2. Methodology

Suppose a population $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_N)$ encompasses of size N independent components. We choose n objects from $\mathbf{Y}$ using simple random sampling without replacement. Consider φ_{iy} and φ_{ix} be the characteristics of the study attributes φ_y and auxiliary attribute φ_x. Deliberate φ_{ij} $(j = 1, 2)$, if i^{th} essentials is chosen and zero otherwise. Let

$$\varphi_j = \left(\frac{a_j}{n}\right) \quad \text{and} \quad P_j = \left(\frac{A_j}{N}\right),$$

signifies the sample and population, where

$$a_j = \sum_{i=1}^{n} \varphi_{ij}, \qquad A_j = \sum_{i=1}^{N} \varphi_{ij}, \qquad \text{as } j = y, x.$$

$$\rho_{\varphi_y\varphi_x} = \left[\frac{S_{\varphi_y\varphi_x}}{s_{\varphi_y}\, s_{\varphi_x}}\right], \qquad S_{\varphi_y\varphi_x} = \left[\frac{1}{N-1}\sum_{i=1}^{N}(\varphi_{iy} - P_y)(\varphi_{ix} - P_x)\right],$$

$$S_{\varphi_y} = \left[\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\varphi_{iy} - P_y)^2}\right], \qquad S_{\varphi_x} = \left[\sqrt{\frac{1}{N}\sum_{i=1}^{N}(\varphi_{ix} - P_x)^2}\right],$$

$$C_{\varphi_j} = \frac{s_{\varphi_j}}{P_j}, \qquad j = y, x,$$

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

$$E(e_0^2) = \lambda C_{\varphi_y}^2, \qquad E(e_1^2) = \lambda C_{\varphi_x}^2, \qquad E(e_0 e_1) = \lambda \rho_{\varphi_y \varphi_x} C_{\varphi_y} C_{\varphi_x},$$

$$\lambda = \left(\frac{1}{n} - \frac{1}{N}\right).$$

3. Existing estimators

In this section, we adopt some existing estimators for population proportion based on simple random sampling, is given by:

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

$$p_{\text{usual}} = \hat{p}_y, \tag{1}$$

$$\text{Var}(p_{\text{usual}}) = \lambda P^2 C_{\varphi_y}^2. \tag{2}$$

(2) The ratio estimator deliberated by [2] with properties are given by:

$$p_{\text{Ratio}} = \hat{p}_y \left(\frac{P_x}{\hat{p}_x}\right), \tag{3}$$

$$\text{Bias}(p_{\text{Ratio}}) \cong \lambda P \left[C_{\varphi_x}^2 - \rho_{\varphi_y \varphi_x} C_{\varphi_y} C_{\varphi_x}\right],$$

$$\text{MSE}(p_{\text{Ratio}}) \cong \lambda P^2 \left[C_{\varphi_y}^2 + C_{\varphi_x}^2 - 2\rho_{\varphi_y \varphi_x} C_{\varphi_y} C_{\varphi_x}\right]. \tag{4}$$

(3) The usual product estimator suggested by [4] properties are given by:

$$p_{\text{Product}} = \hat{p}_y \left(\frac{\hat{p}_x}{P_x}\right) \tag{5}$$

$$\text{Bias}(p_{\text{Product}}) \cong \lambda P \left(\rho_{\varphi_y \varphi_x} C_{\varphi_y} C_{\varphi_x} - \frac{1}{2} C_{\varphi_x}^2\right)$$

$$\text{MSE}(p_{\text{Product}}) \cong \lambda P^2 \left[C_{\varphi_y}^2 + C_{\varphi_x}^2 + 2\rho_{\varphi_y \varphi_x} C_{\varphi_y} C_{\varphi_x}\right] \tag{6}$$

(4) The usual regression estimator with variance are given by:

$$p_{\text{Regression}} = \hat{p}_y + b(P_x - \hat{p}_x) \tag{7}$$

$$\text{Var}(p_{\text{Regression}}) = \text{MSE}(p_{\text{Regression}}) = \lambda P^2 C_{\varphi_y}^2 \left(1 - \rho_{\varphi_y \varphi_x}^2\right) \tag{8}$$

(5) The authors in [3] recommended the following ratio and product type estimators along with properties as:

$$p_{\text{Bahl}} = \hat{p}_y \exp\left(\frac{P_x - \hat{p}_x}{P_x + \hat{p}_x}\right) \tag{9}$$

and

$$p_{\text{BahlP}} = \hat{p}_y \exp\left(\frac{\hat{p}_x - P_x}{\hat{p}_x + P_x}\right) \tag{10}$$

$$\text{Bias}(p_{\text{BahlR}}) \cong \lambda P_y \left(\frac{3}{8} C^2_{\varphi_x} - \frac{1}{2} \rho_{\varphi_y \varphi_x} C_{\varphi_y} C_{\varphi_x}\right) \tag{11}$$

$$\text{MSE}(p_{\text{BahlR}}) = \lambda P^2 \left(C^2_{\varphi_y} + \frac{1}{4} C^2_{\varphi_x} - \rho_{\varphi_y \varphi_x} C_{\varphi_y} C_{\varphi_x}\right) \tag{12}$$

$$\text{Bias}(p_{\text{BahlP}}) \cong \lambda P_y \left(\rho_{\varphi_y \varphi_x} C_{\varphi_y} C_{\varphi_x} - \frac{1}{8} C^2_{\varphi_x}\right) \tag{13}$$

$$\text{MSE}(p_{\text{BahlP}}) \cong \lambda P^2 \left[C^2_{\varphi_y} + \frac{1}{4} C^2_{\varphi_x} + \rho_{\varphi_y \varphi_x} C_{\varphi_y} C_{\varphi_x}\right] \tag{14}$$

(6) The improved difference estimators for population proportion established by [5] and [6] are given by:

$$P_{D1} = G_1 P_y + Q_2 (P - \hat{P}_x) \tag{15}$$

$$P_{D2} = G_2 P_y + G_4 (P - \hat{P}_x) \left(\frac{\hat{p}_x}{P}\right) \tag{16}$$

Where G_1, G_2, G_3 and G_4 are constants.

$$G_2 = \frac{P_y}{P_x} \left[\frac{\rho_{\varphi_y \varphi_x} C_{\varphi_y} C_{\varphi_x}}{C^2_{\varphi_x}}\right]$$

$$G_3 = \frac{1 - \lambda C^2_{\varphi_y}}{1 - \lambda C^2_{\varphi_x} + 2\lambda C_{\varphi_y} C_{\varphi_x}}$$

$$G_4 = \frac{P_y}{P_x} \left[1 + \left(\frac{\rho_{\varphi_y \varphi_x} C_{\varphi_y} C_{\varphi_x}}{C_{\varphi_x}}\right) - 2\right]$$

The properties of P_{D1} and P_{D2} are given by:

$$\text{Bias}(P_{D1}) = \frac{\lambda P_y C^2_{\varphi_y} \left(1 - \rho^2_{\varphi_y \varphi_x}\right)}{1 + \lambda C^2_{\varphi_y} \left(1 - \rho^2_{\varphi_y \varphi_x}\right)} \tag{17}$$

$$\text{Bias}(P_{D2}) = \lambda P \left[\frac{\left[5 C^2_{\varphi_y} (1 - \rho^2_{\varphi_y \varphi_x}) - \lambda C^2_{\varphi_x} S^2_{\varphi_y} + C_{\varphi_y} C_{\varphi_x} (\rho_{\varphi_y \varphi_x} - 1)(1 - \lambda C^2_{\varphi_x})\right]}{1 - \lambda C_{\varphi_y} \rho_{\varphi_y \varphi_x}}\right] \tag{18}$$

$$\mathrm{MSE}(P_{D1}) = \frac{\lambda P^2 C_{\varphi_y}^2\left(1-\rho_{\varphi_y\varphi_x}^2\right)}{1+\lambda C_{\varphi_y}^2\left(1-\rho_{\varphi_y\varphi_x}^2\right)} \tag{19}$$

$$\mathrm{MSE}(P_{D2}) = \frac{\lambda P^2(1-\lambda C_{\varphi_x}^2)C_{\varphi_y}^2(1-\rho_{\varphi_y\varphi_x}^2)}{(1-\lambda C_{\varphi_x}^2)+4\lambda C_{\varphi_y}(1-\rho_{\varphi_y\varphi_x}^2)} \tag{20}$$

4. Recommended generalized class of estimators

Precise estimation of population proportion is a fundamental task in statistical survey methodology, supporting important decisions in areas ranging from public health surveillance and environmental risk management to radiological safety assessment. While esteemed, classical estimators are often inefficient and delicate in a variety of practical settings where sample availability is sparse, data collection is expensive or measurement is intrinsically risky. In many practical applications, information in the form of categorical or binary attributes is available and naturally associated with the response. Nevertheless, existing estimation frameworks tend to neglect these auxiliary features in the form of ratio or simple regression estimators, and consequently complex dependent structures between them that lead to sub-optimal estimator performance. The radiation data science domain represents a ideal example of this problem, where direct radiation science measurements can be expensive to execute due to logistical reasons (sensor deployment, transport etc.), financial constraints and safety concerns but in bound information coming from auxiliary indicators i.e., spatial closeness with sources of contamination, land use typologies, historical source of radiation exposure readings are available as indirect yet informative proxies. Utilizing these supplemental dimensions to improve estimates of population proportions is a critical but underexplored methodological need.

At the same time, however, there exists a conspicuous gap in the literature: so few specialized and theoretically rigorous estimators have been developed that can be used universally to exploit categorical auxiliary information when estimating population proportions; as well remains an complete insufficiency of empirical verification accomplished using real-world radiation data. Moreover, many existing methods are not robust to attribute misclassification, measurement errors and heterogeneous structure of correlations; thus it is challenging to use them in more complex realistic cases. This research stems from the persistent necessity of developing flexible, analytically tractable and empirically justified classes of estimators that naturally combine auxiliary attribute to study attribute that greatly improve on estimation performances. By filling this methodological gap and inserting the mechanism indeed in operational ties of radiation exposure assessment, the research aims at a substantial contribution to the theory of survey sampling as well as to its seminal arena of practical applications in environmental health and radiological risk governance.

By taking motivation from [1], we developed a generalized class of estimators for population proportion using auxiliary attributes. The authors in [25] developed the following improved estimator:

$$p_{\text{Rao}} = P_y - D_1\left(\hat{p}_x - 1\right) \tag{21}$$

The author in [26] developed the following estimator, which is given by:

$$P_{\text{Singh}} = \hat{P}_y\left(\frac{P_x}{\hat{p}_x}\right) - \alpha\left(\frac{\hat{p}_x}{P_x} - 1\right) \tag{22}$$

Where α is constant. Ref. [3] developed the following exponential ratio type estimators, which is given by:

$$P_{\text{BahlR}} = \hat{p}_y \exp\left(\frac{P_x - \hat{p}_x}{P_x + \hat{p}_x}\right) \tag{23}$$

Using (21) and (22) we modify (23), which is given by:

$$P_{\text{BahlR1}} = P_y \exp\left(\frac{P_x - \hat{p}_x}{P_x + \hat{p}_x}\right) - \alpha\left(\frac{\hat{p}_x}{P_x} - 1\right) \tag{24}$$

On the lines of [27], we have written by:

$$\hat{P}_{\text{BahlR2}} = \hat{p}_y \exp\left(\frac{aP_x + b\hat{p}_x}{a(a+b) + b(1-a)(aP_x + b)} - 1\right). \tag{25}$$

Ref. [5] developed the following estimator:

$$\hat{P}_{\text{Rao-D}} = D_1\hat{p}_y + D_2(P_x - \hat{p}_x) \tag{26}$$

Replace $\hat{P}_{\text{Rao}}$ given in (26) instead of $\hat{p}_y$ in (23) and also replace D_1 and D_2 is given by:

$$\hat{P}_{\text{Prop}} = \left[T_1\hat{p}_y + T_2(P_x - \hat{p}_x)\right]\exp\left(\frac{aP_x + b}{a(\alpha+b) + b(1-\alpha)(aP_x + b)} - 1\right) \tag{27}$$

Table 1. Some members of the recommended class of estimators

Estimators	α	a	b
$\hat{P}_{\text{Prop1}} = \left[T_1\hat{p}_y + T_2(P_x - \hat{p}_x)\right]\exp\left(\frac{P_x+P_{\varphi_x}}{P_x+P_{\varphi_x}} - 1\right)$	1	1	P_{φ_x}
$\hat{P}_{\text{Prop2}} = \left[T_1\hat{p}_y + T_2(P_x - \hat{p}_x)\right]\exp\left(\frac{P_x+C_{\varphi_x}}{P_x+C_{\varphi_x}} - 1\right)$	1	1	C_{φ_x}
$\hat{P}_{\text{Prop3}} = \left[T_1\hat{p}_y + T_2(P_x - \hat{p}_x)\right]\exp\left(\frac{P_x+\beta_{\varphi_x}}{P_x+\beta_{\varphi_x}} - 1\right)$	1	1	β_{φ_x}
$\hat{P}_{\text{Prop4}} = \left[T_1\hat{p}_y + T_2(P_x - \hat{p}_x)\right]\exp\left(\frac{\theta_{\varphi_x}P_x+C_{\varphi_x}}{\theta_{\varphi_x}P_x+C_{\varphi_x}} - 1\right)$	1	θ_{φ_x}	C_{φ_x}
$\hat{P}_{\text{Prop5}} = \left[T_1\hat{p}_y + T_2(P_x - \hat{p}_x)\right]\exp\left(\frac{C_{\varphi_x}P_x+\beta_{\varphi_x}}{C_{\varphi_x}P_x+\beta_{\varphi_x}} - 1\right)$	1	C_{φ_x}	β_{φ_x}

Solving (27) up to the first order of approximation, we have:

$$\hat{P}_{\text{Prop}} = P_y - P_y + P_yT_1 + P_ye_0T_1 - P_y\theta e_1T_1 - P_y\theta e_0e_1T_1 + \frac{P_y\theta^2e_1^2}{2}T_1 - P_xe_1T_2 + P_xe_0e_1T_2 \tag{28}$$

Where

$$\theta = \alpha a$$

$$\theta = \frac{aP_x}{\alpha P_x + b}$$

Using (28), we obtain the bias and mean squared error up to the first order of approximation, we have:

$$\text{Bias}(P_{\text{Prop}}) = -P_y + P_y T_1 \lambda C_{\varphi_x}^2 + T_1 \left(P_y + \frac{3P_y \theta^2 \lambda C_{\varphi_x}^2}{2} - P_y \theta \rho_{\varphi_y \varphi_x} C_{\varphi_y} C_{\varphi_x} \right) \tag{29}$$

And

$$\begin{aligned} \text{MSE}(P_{\text{Prop}}) = & P_y^2 + P_x^2 T_2 \left(-2P_y\theta + P_x T_2\right) \lambda C_{\varphi_x}^2 \\ & + P_y T_1 \left[-2P_y + \theta\left(-3P_y\theta + 4P_x T_2\right)\right] \lambda C_{\varphi_x}^2 \\ & + 2(P_y\theta - P_x T_2) P_{\varphi_y\varphi_x} C_{\varphi_y} C_{\varphi_x} \\ & + P_y^2 T_1^2 \left(1 + 4\theta^2 \lambda C_{\varphi_x}^2 - 4\theta \rho_{\varphi_y\varphi_x} C_{\varphi_y} C_{\varphi_x} + \lambda C_{\varphi_x}^2\right) \end{aligned} \tag{30}$$

5. Numerical study

In this section, we used some radiation data sets to check the efficiency of an estimators. 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: [20]]

$$P_1 = \text{Proportion } \phi_{i1} < 1 \text{ for } (Y \leq 106) \text{ and } \phi_{i1} > 1 \text{ for } (Y > 106)$$

$$P_2 = \text{Proportion } \phi_{i2} < 1 \text{ for } (X \leq 106.1) \text{ and } \phi_{i2} > 1 \text{ for } (X > 106.1)$$

$$N = 21,\ n = 6,\ P_1 = 0.3809524,\ P_2 = 0.2857143,\ \lambda = 0.1190476,$$

$$\rho_{\phi_1\phi_2} = -0.4961389,\ C_{\phi_1} = 1.306235,\ C_{\phi_2} = 1.620185$$

Population–II: [Source: [21]]

$$P_1 = \text{Proportion } \phi_{i1} < 1 \text{ for } (Y \leq 122) \text{ and } \phi_{i1} > 1 \text{ for } (Y > 122)$$

$$P_2 = \text{Proportion } \phi_{i2} < 1 \text{ for } (X \leq 121) \text{ and } \phi_{i2} > 1 \text{ for } (X > 121)$$

$$N = 20,\ n = 5,\ P_1 = 0.25,\ P_2 = 0.25,\ \lambda = 0.15,$$

$$\rho_{\phi_1\phi_2} = -0.06666667,\ C_{\phi_1} = 1.777047,\ C_{\phi_2} = 1.777047$$

Population–III: [Source: [22]]

$$P_1 = \text{Proportion } \phi_{i1} < 1 \text{ for } (Y \leq 1.920) \text{ and } \phi_{i1} > 1 \text{ for } (Y > 0.7389000)$$

$$P_2 = \text{Proportion } \phi_{i2} < 1 \text{ for } (X \leq 0.01802) \text{ and } \phi_{i2} > 1 \text{ for } (X > 2.22600)$$

$$N = 21,\ n = 6,\ P_1 = 0.4761905,\ P_2 = 0.5238095,\ \lambda = 0.1190476,$$

$$\rho_{\phi_1\phi_2} = 0.5272727,\ C_{\phi_1} = 1.074709,\ C_{\phi_2} = 0.9770084$$

Table 2. Numerical result of MSEs of all estimators using radiation data science

Estimators	**I**	**II**	**III**
$\widehat{P}_U$	0.02947846	0.02960526	0.03117914
$\widehat{P}_{Ratio}$	0.1111111	0.06315789	0.02705628
$\widehat{P}_{Product}$	0.03854875	0.05526316	0.08683777
$\widehat{P}_{Reg}$	0.02222222	0.02947368	0.02251082
$\widehat{P}_{ER}$	0.05895692	0.03898026	0.02267574
$\widehat{P}_{EP}$	0.02267574	0.03503289	0.05256648
$\widehat{P}_{D1}$	0.0192713	0.02002861	0.02047792
$\widehat{P}_{D2}$	0.02106568	0.01554519	0.01969853
$\widehat{P}_{Prop1}$	0.004951806	0.007112226	0.003360327
$\widehat{P}_{Prop2}$	0.006193714	0.004185456	0.01091065
$\widehat{P}_{Prop3}$	0.008560236	0.005898659	0.01226231
$\widehat{P}_{Prop4}$	0.005874292	0.004809526	0.01191328
$\widehat{P}_{Prop5}$	0.00919856	0.006654765	0.01226192

Table 3. Numerical result of PREs of all estimators using radiation data science

Estimators	**I**	**II**	**III**
$\widehat{P}_U$	100	100	100
$\widehat{P}_{Ratio}$	26.53061	46.875	115.2381
$\widehat{P}_{Product}$	76.47059	53.57143	35.90504
$\widehat{P}_{Reg}$	132.6531	100.4464	138.5073
$\widehat{P}_{ER}$	50	75.94937	137.5
$\widehat{P}_{EP}$	130	84.50704	59.31373
$\widehat{P}_{D1}$	152.9656	147.8148	152.2573
$\widehat{P}_{D2}$	139.936	190.4464	158.2816
$\widehat{P}_{Prop1}$	595.3072	416.2587	927.8601
$\widehat{P}_{Prop2}$	475.9415	707.3366	285.7679
$\widehat{P}_{Prop3}$	344.365	501.8982	254.2681
$\widehat{P}_{Prop4}$	501.8215	615.5547	261.7175
$\widehat{P}_{Prop5}$	320.4682	444.8732	254.2762

6. Simulation study

To evaluate the precision of a variance estimator by generating synthetic data from a bivariate normal distribution and analyzing the estimator's bias, consistency, and efficiency. We have created three groups of 1,000 people each, based on a bivariate normal distribution, that each have their own covariance matrices. We select a sample of size $n = 200$ from each group. The following are the means and covariance matrices for the population.

Population–I:

$$\mu_1 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

and

$$\Sigma_1 = \begin{bmatrix} 2 & 3 \\ 3 & 5 \end{bmatrix}$$

$$\rho_{yx} = 0.7653333$$

Population–II:

$$\mu_2 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

and

$$\Sigma_2 = \begin{bmatrix} 2 & 3 \\ 3 & 10 \end{bmatrix}$$

$$\rho_{yx} = 0.4186667$$

Population–III:

$$\mu_3 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$$

and

$$\Sigma_3 = \begin{bmatrix} 2 & -2 \\ -2 & 4 \end{bmatrix}$$

$$\rho_{yx} = -0.3013333$$

Table 4. Numerical result of MSEs using simulation analysis

Estimators	**I**	**II**	**III**
$\widehat{P}_U$	0.001000857	0.001000677	0.001000857
$\widehat{P}_{Ratio}$	0.0003661493	0.00108807	0.003183049
$\widehat{P}_{Product}$	0.003702371	0.003098772	0.0009699738
$\widehat{P}_{Reg}$	0.0003276942	0.0007694398	0.0007162801
$\widehat{P}_{ER}$	0.0004251522	0.0007711871	0.001823039
$\widehat{P}_{EP}$	0.002093263	0.001776538	0.0007165017
$\widehat{P}_{D1}$	0.0003272753	0.0007671614	0.0007141838
$\widehat{P}_{D2}$	0.0003273154	0.0007674256	0.0007143945
$\widehat{P}_{Prop1}$	0.0002084775	0.0003130525	0.0003688099
$\widehat{P}_{Prop2}$	0.0002178674	0.000386804	0.0003587129
$\widehat{P}_{Prop3}$	0.0002482573	0.0004414447	0.0004126933
$\widehat{P}_{Prop4}$	0.0002308294	0.0003992351	0.0003586449
$\widehat{P}_{Prop5}$	0.0002482573	0.0004414495	0.0004127213

Table 5. Numerical result of PREs using simulation analysis

Estimators	**I**	**II**	**III**
$\widehat{P}_U$	100	100	100
$\widehat{P}_{Ratio}$	273.3467	91.96808	31.44334
$\widehat{P}_{Product}$	27.03286	32.29269	103.1839
$\widehat{P}_{Reg}$	305.424	130.0526	139.7298
$\widehat{P}_{ER}$	235.4114	129.758	54.90046
$\widehat{P}_{EP}$	47.81323	56.32733	139.6866
$\widehat{P}_{D1}$	305.8149	130.4389	140.1399
$\widehat{P}_{D2}$	305.7775	130.394	140.0986
$\widehat{P}_{Prop1}$	480.079	319.6514	271.3747
$\widehat{P}_{Prop2}$	459.3881	258.7038	279.0133
$\widehat{P}_{Prop3}$	403.1531	226.6822	242.5183
$\widehat{P}_{Prop4}$	433.5916	250.6485	279.0663
$\widehat{P}_{Prop5}$	403.1531	226.6798	242.5018

7. Discussion

In this paper, we developed a new generalized class of estimators for population proportion under simple random sampling by utilizing auxiliary attribute information. The theoretical basis and the detailed derivation of bias and MSE under first-order approximation establish rigorous analytical support for the performance analysis and parameter adaptation of the estimator. The simulation study performed in this paper provides strong empirical evidence on the efficiency of the estimator compared with standard approaches such as classical ratio-type estimators.

Crucially, the empirical application of our methodology to radiation science data also supports the effectiveness of using this estimator in practice. Radiation data are, by definition, very challenging: direct measurements are typically limited due to safety issues and/or logistical and financial constraints, while complementary categorical covariates such as the distance to areas of contamination or land use classification are essential surrogates. In this setting, the proposed estimator demonstrates much better efficiency and validity as compared to conventional estimators.

Such efficiency gains are particularly relevant in the context of radiological risk assessment, where precise knowledge about population proportion would be pivotal for informed policy decisions, regulatory control, and public health actions. Members of the recommended class of estimators are given in Table 1. Numerical results of the MSEs and PREs using radiation data science are given in Tables 2 and 3. Using simulation analysis, results of MSEs and PREs of all considered estimators are given in Tables 4 and 5.

The more efficient performance of the proposed estimators is due to their ability to make better use of auxiliary information that was previously under-utilized or used in a restricted way. By appropriately incorporating such knowledge, the estimator realizes substantial variance reduction without introducing sizable bias, even when the auxiliary data are unreliable or noisy. This property confirms its strength and practical applicability in field scenarios where auxiliary features are naturally indirect predictors and may also be subject to misclassification.

Additional sensitivity analyses showed that this estimator remains more efficient over a wide range of correlation strengths between study variables and auxiliary variables, supporting its applicability to generic empirical contexts beyond radiation data. This makes the estimator a flexible tool applicable to a variety of fields, such as epidemiology, environmental sciences, and social sciences.

Nonetheless, certain limitations merit consideration. The present methodology is established primarily under the assumption of simple random sampling without replacement, and its generalization to other complex survey designs, such as stratified or cluster sampling, would be a topic for future research. Further investigation of the behavior of the estimator under continuous and mixed-type auxiliary variables, measurement errors, and incomplete data is needed to enhance its applicability.

In conclusion, simulation and radiation data science confirm that the recommended generalized class of estimators presents a methodological advance by combining theoretical elegance with empirical effectiveness. This work consequently contributes to the statistics of population proportion estimation, with direct implications for radiation data science and other applied areas that depend on attribute information to improve inferential accuracy.

8. Conclusion

In this manuscript, we proposed a new generalized class of estimators for statistical inference of population proportion estimation which possesses wide-ranging applications and effortlessly integrates auxiliary attribute information in the form of qualitative/categorical or quantitative/binary type variables.

The developed theoretical framework with detailed derivations of bias and mean squared error under first-order approximation facilitates the optimization of parameters for the proposed class of estimators, leading to significant improvement in both estimation accuracy and efficiency. Through extensive simulation analysis, we repeatedly observed that the proposed class of estimators dominates the classical estimators in bias reduction and mean square error across a wide range of sampling designs.

In addition, empirical application to real radiation science data validates the practical usefulness of the estimator with significant improvements in accuracy and robustness under challenging realistic scenarios involving auxiliary attributes.

The capacity of the estimator to naturally accommodate auxiliary variables dramatically expands its scope of application, resulting in a valuable tool for several important areas such as environmental monitoring, radiological risk assessment, public health investigations, and disciplines constrained by financial or safety limitations.

Although this paper focuses specifically on population proportion under simple random sampling using auxiliary attributes, the present study improves the boundary of survey sampling by recommending a theoretically well-founded, empirically demonstrated, and practically useful estimator for population proportion. The general applicability and improved accuracy make it an important tool for statisticians and researchers working on data-driven decision making in many applied fields.

Conflicts of Interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

[1] S. Khan, M. Farooq, S. Ahmad, S. Khan, Improved Estimator for Estimation of Population Mean Using Predictive Approach Under PPS Sampling, VFAST Trans. Math. 12 (2024), 01–16. https://doi.org/10.21015/vtm.v12i2.1942.

[2] W.G. Cochran, The Estimation of the Yields of Cereal Experiments by Sampling for the Ratio of Grain to Total Produce, J. Agric. Sci. 30 (1940), 262–275. https://doi.org/10.1017/s0021859600048012.

[3] S. Bahl, R. Tuteja, Ratio and Product Type Exponential Estimators, J. Inf. Optim. Sci. 12 (1991), 159–164. https://doi.org/10.1080/02522667.1991.10699058.

[4] M.N. Murthy, Product Method of Estimation, Sankhya: Indian J. Stat. Ser. A (1964), 69–74.

[5] T.J. Rao, On Certail Methods of Improving Ration and Regression Estimators, Commun. Stat. - Theory Methods 20 (1991), 3325–3340. https://doi.org/10.1080/03610929108830705.

[6] S. Gupta, J. Shabbir, S. Ahmad, Estimation of Median in Two-Phase Sampling Using Two Auxiliary Variables, Commun. Stat. - Theory Methods 37 (2008), 1815–1822. https://doi.org/10.1080/03610920701826476.

[7] S.M. Alghamdi, S. Ahmad, S.M. Almarzouki, B. Aloraini, M.M. Badr, et al., Constructing a New Estimator for Estimating Population Mean Utilizing Auxiliary Information in Probability Proportional to Size Sampling, Alex. Eng. J. 110 (2025), 506–511. https://doi.org/10.1016/j.aej.2024.10.029.

[8] A. Abd-Elfattah, E. El-Sherpieny, S. Mohamed, O. Abdou, Improvement in Estimating the Population Mean in Simple Random Sampling Using Information on Auxiliary Attribute, Appl. Math. Comput. 215 (2010), 4198–4202. https://doi.org/10.1016/j.amc.2009.12.041.

[9] S. Bhushan, A. Kumar, Improved Estimation of Population Mean in Simple Random Sampling Using Attribute, Thailand Statistician 22 (2024), 374–389.

[10] S. Bhushan, A. Kumar, Evaluating the Performance of Logarithmic Type Estimators Using Auxiliary Attribute, Life Cycle Reliab. Saf. Eng. 12 (2023), 285–292. https://doi.org/10.1007/s41872-023-00236-4.

[11] S. Ahmad, J. Shabbir, Proportion Estimation Using Enhanced Class of Estimators Under Simple Random Sampling: Application with Real Data Sets and Simulation, Measurement: Interdiscip. Res. Perspect. 23 (2024), 297–309. https://doi.org/10.1080/15366367.2024.2346426.

[12] S. Bhushan, A. Kumar, S. Kumar, Efficient Classes of Estimators for Population Variance Using Attribute, Int. J. Math. Oper. Res. 22 (2022), 74–92. https://doi.org/10.1504/ijmor.2022.123124.

[13] M. Saini, B.R. Jitendrakumar, A. Kumar, Optimum Estimator in Simple Random Sampling Using Two Auxiliary Attributes with Application in Agriculture, Fisheries and Education Sectors, MethodsX 9 (2022), 101915. https://doi.org/10.1016/j.mex.2022.101915.

[14] H. Semary, S. Ahmad, I. Elbatal, C. Chesneau, M. Elgarhy, et al., New Comprehensive Class of Estimators for Population Proportion Using Auxiliary Attribute: Simulation and an Application, Alex. Eng. J. 99 (2024), 130–136. https://doi.org/10.1016/j.aej.2024.04.065.

[15] H.P. Singh, R.S. Solanki, Improved Estimation of Population Mean in Simple Random Sampling Using Information on Auxiliary Attribute, Appl. Math. Comput. 218 (2012), 7798–7812. https://doi.org/10.1016/j.amc.2012.01.047.

[16] T. Zaman, E. Toksoy, Improvement in Estimating the Population Mean in Simple Random Sampling Using Information on Two Auxiliary Attributes and Numerical Application in Agricultural Engineering, Fresenius Environ. Bull. 28 (2019), 4584–4590.

[17] S. Bhushan, A. Kumar, R. Gupta, Novel Logarithmic Estimator Using Two Auxiliary Attributes, Int. J. Agric. Stat. Sci. 18 (2022), 367.

[18] E. Zamanzade, X. Wang, Estimation of Population Proportion for Judgment Post-Stratification, Comput. Stat. Data Anal. 112 (2017), 257–269. https://doi.org/10.1016/j.csda.2017.03.016.

[19] E. Zamanzade, X. Wang, Proportion Estimation in Ranked Set Sampling in the Presence of Tie Information, Comput. Stat. 33 (2018), 1349–1366. https://doi.org/10.1007/s00180-018-0807-x.

[20] A.R. El-Saeed, S. Ahmad, B. Aloraini, Robust Estimator for Estimation of Population Mean Under PPS Sampling: Application to Radiation Data, J. Radiat. Res. Appl. Sci. 18 (2025), 101384. https://doi.org/10.1016/j.jrras.2025.101384.

[21] B. Aloraini, L. Diab, S. Ahmad, Regression Cum Exponential Type Estimator for Population Proportion Using Auxiliary Attributes: Application on Radiation Data and Simulation Study, J. Radiat. Res. Appl. Sci. 18 (2025), 101479. https://doi.org/10.1016/j.jrras.2025.101479.

[22] G. Alomani, M. Kayid, S. Ahmad, Application to Radiation Data Sets by Suggesting an Improved Mean Estimator Under Probability Proportional to Size Sampling, J. Radiat. Res. Appl. Sci. 18 (2025), 101354. https://doi.org/10.1016/j.jrras.2025.101354.

[23] T. Zaman, New Family of Estimators Using Two Auxiliary Attributes, Int. J. Adv. Res. Eng. Manage. 4 (2018), 11–16.

[24] T. Zaman, C. Kadilar, Novel Family of Exponential Estimators Using Information of Auxiliary Attribute, J. Stat. Manag. Syst. 22 (2019), 1499–1509. https://doi.org/10.1080/09720510.2019.1621488.

[25] B. Prasad, Some Unbiased Estimators Versus Mean Per Unit and Ratio Estimators in Finite Population Sample Surveys, Commun. Stat. - Theory Methods 15 (1986), 3647–3657. https://doi.org/10.1080/03610928608829338.

[26] R. Singh, H.P. Singh, M. Espejo, The Efficiency of an Alternative to Ratio Estimator Under a Super Population Model, J. Stat. Plan. Inference 71 (1998), 287–301. https://doi.org/10.1016/s0378-3758(98)00009-3.

[27] M. Khoshnevisan, R. Singh, P. Chauhan, N. Sawan, F. Smarandache, A General Family of Estimators for Estimating Population Mean Using Known Value of Some Population Parameter(s), Far East J. Theor. Stat. 22 (2007), 181–191.