Journal of Informatics and Mathematical Sciences
Vol. 17, No. 4, pp. 351–364, 2025
ISSN 0975-5748 (online); 0974-875X (print)
Published by RGN Publications
DOI: 10.26713/jims.v17i4.3424

http://www.rgnpublications.com

Research Article

Enhanced Survey Estimation Using Calibration under Scrambled Response and Measurement Error

Ridhi Agarwal* and Yash Prakash

AcadAlly, Avalon Labs Pvt. Ltd., A-91, First Floor, Nagpal Towers, Okhla Phase II, New Delhi 110020, India
***Corresponding author:** ridhi@acadally.com

Received: October 15, 2025 **Revised:** November 27, 2025
Accepted: December 8, 2025 **Published:** December 30, 2025

Abstract. This study addresses the challenges of collecting accurate data on sensitive issues by applying various Scrambled Response Techniques combined with various calibration estimators under measurement error. A simulation study using real data evaluates the performance of the proposed estimators both with and without measurement error. Results show that the proposed method consistently outperforms traditional Scrambled Response Technique, demonstrating greater efficiency and reliability in handling sensitive survey data.

Keywords. Auxiliary information, Calibration estimators, Mean estimation, Measurement error, Scrambled Response Technique (SRT)

Mathematics Subject Classification (2020). 62D05

1. Introduction

Gathering reliable data on sensitive or socially stigmatized issues such as drug use, length of time living with AIDS, sexual practices, tax evasion, or domestic violence can be extremely difficult. When direct questioning methods are used, respondents may either withhold information or give inaccurate answers due to fear, embarrassment, or social pressure. This can lead to serious flaws in the data and undermine the credibility of the research findings. To address this problem and encourage greater honesty and participation, researchers have developed various indirect questioning techniques. These approaches are designed to safeguard respondent anonymity and privacy, thereby increasing the likelihood of obtaining truthful and complete responses in surveys dealing with sensitive topics.

Therefore, in such situations, methods that ensure respondent anonymity offer an effective solution. Two widely adopted approaches that provide this protection are the *Randomized Response Technique* (RRT) and the *Scrambled Response Technique* (SRT). The concept of randomized response was first introduced by Warner [18]. Later, the scrambled response technique was proposed by Pollock and Bek [10]. Since then, several researchers have made significant contributions to this field, including Eichhorn and Hayre [5], Saha [14], Diana and Perri [3,4], Perri and Diana [9], and as well as Priyanka *et al*. [12]. These works have collectively enriched the literature on indirect questioning methods aimed at improving data reliability in sensitive surveys.

Surveys focusing on key socio-economic issues such as unemployment, wages, and various labor and health-related factors are routinely conducted by government agencies over time. However, these surveys are often prone to measurement errors, which can affect the reliability and validity of the collected data. One common source of such error is sampling error, which occurs when data is gathered from a subset of the population rather than the entire group, leading to potential inaccuracies in the results.

To address these challenges, several renowned researchers have proposed effective estimation techniques that account for measurement errors in assessing population characteristics. Pioneering work in this area has been carried out by Mahalanobis [8], Deming [1], Raj [13], Sukhatme *et al*. [17], Sarndal *et al*. [15], Gregoire and Salas [6], Vishwakarma *et al*. [16] and Priyanka *et al*. [11]. Their contributions have played a significant role in improving the accuracy and effectiveness of survey based research under conditions of measurement error.

Building on the contributions of renowned researchers in the field, this study focuses on addressing the challenges associated with sensitive survey variables by employing various *Scrambled Response Techniques* (SRT) alongside a range of various calibration estimators in the presence of measurement errors. To evaluate the effectiveness of the proposed methodology, a simulation study using real-life data has been carried out. This study compares the performance of the estimators under two conditions, one involving measurement error and the other assuming no measurement error.

The simulation results demonstrate that the proposed approach consistently delivers higher efficiency than the standard SRT methods with measurement error. These findings emphasize the robustness and practical applicability of the model in effectively handling sensitive survey data while maintaining respondent privacy and improving data accuracy.

2. Survey Setup and Notation

Consider a finite population $U = (U_1, U_2, \dots, U_N)$ consisting of N distinct and identifiable units. Let y denote the sensitive variable of interest, and x represent a non-sensitive auxiliary variable. The population means of y and x are denoted by $\bar{Y}$ and $\bar{X}$, respectively. Our goal is to estimate the mean $\bar{Y}$ while accounting for the presence of measurement errors.

A sample s_n of size n is drawn from the population according to a sampling design d, where each unit U_i has an inclusion probability $\pi_i = P(U_i \in s_n)$, and each pair of units (U_i, U_j) has a joint inclusion probability $\pi_{ij} = P(U_i, U_j \in s_n)$. Define $\Delta_{ij} = \pi_{ij} - \pi_i \pi_j$ as the covariance between inclusion indicators.

Within this sampling framework, we intend to apply the scrambled response technique to effectively address the sensitivity of the study variable.

3. Scrambled Response Technique under Measurement Error

Following the methodology introduced by Priyanka *et al*. [12], we employ various *Scrambled Response Techniques* (SRT) to estimate the population mean of a sensitive variable. In this work, their original framework is extended by incorporating various calibration estimators under the presence of measurement errors with SRT.

To preserve respondent privacy in surveys involving sensitive topics, the true values of sensitive variables are deliberately perturbed using scrambling variables-an approach known as the SRT. Building on established literature in this area, we present a model in which the sensitive variable y is transformed into a scrambled response variable z as follows:

SRT-I

$$z = y\left[O_1 + \frac{O_2}{y}\right], \tag{1}$$

where O_1 and O_2 are scrambling variables that may follow any suitable distribution. Under this transformation, the population mean $\bar{Y}$ of the sensitive variable can be recovered from the observed mean $\bar{Z}$ using:

$$\bar{Y}_{ts} = \frac{\bar{Z} - \bar{O}_2}{\bar{O}_1}. \tag{2}$$

Additionally, the correlation between the scrambled response variable z and the auxiliary variable x is given by:

$$\rho_{zx} = \frac{\rho_{yx}\sigma_y \bar{O}_1}{\sqrt{\sigma_y^2(\sigma_{O_1}^2 + \bar{O}_1^2) + \sigma_{O_1}^2 \bar{Y}^2 + \sigma_{O_2}^2}}.$$

The general model in eq. (1) encompasses several well-known scrambling models as special cases, depending on the choice of scrambling variables. Two notable cases are outlined below.

Case 1 (SRT-II): If we set $O_1 = 1$ in eq. (1), the model simplifies to the additive scrambling model proposed by Pollock and Bek [10]:

$$z_a = y + O_2. \tag{3}$$

The corresponding estimate of the population mean becomes:

$$\bar{Y}_a = \bar{Z}_a - \bar{O}_2 \tag{4}$$

with

$$\rho_{zx} = \frac{\rho_{yx}\sigma_y}{\sqrt{\sigma_y^2 + \sigma_{O_2}^2}}, \quad C_z^2 = \frac{\sigma_y^2 + \sigma_{O_2}^2}{(\bar{Y} + \bar{O}_2)^2}$$

Case 2 (SRT-III): Setting $O_2 = 0$ in eq. (1), we obtain the multiplicative model discussed by Eichhorn and Hayre [5]:

$$z_m = yO_1 \tag{5}$$

and the estimator of the population mean is:

$$\bar{Y}_m = \frac{\bar{Z}_m}{\bar{O}_1} \tag{6}$$

with

$$\rho_{zx} = \frac{\rho_{yx}\bar{O}_1 C_y}{\sqrt{\sigma_y^2(\sigma_{O_1}^2 + \bar{O}_1^2) + \bar{Y}^2\sigma_{O_1}^2}}, \quad C_z^2 = \frac{\sqrt{\sigma_y^2(\sigma_{O_1}^2 + \bar{O}_1^2) + \bar{Y}^2\sigma_{O_1}^2}}{(\bar{Y}\bar{O}_1)^2}.$$

SRT-IV: We consider the model proposed by Saha [14], in which the sensitive variable y is perturbed into a scrambled variable z_s as follows:

$$z_s = O_1(y + O_2). \tag{7}$$

Based on this scrambling mechanism, the estimator for the population mean is given by:

$$\bar{Y}_s = \frac{\bar{Z}_s - \bar{O}_1\bar{O}_2}{\bar{O}_1} \tag{8}$$

The associated statistical expressions are:

$$\rho_{zx} = \frac{\rho_{yx}\bar{O}_1 S_y}{\sqrt{S_{O_1}^2(S_y^2 + \bar{Y}^2 + S_{O_2}^2 + \bar{O}_2^2) + \bar{O}_1^2(S_y^2 + S_{O_2}^2)}}$$

$$C_z^2 = \frac{S_{O_1}^2(S_y^2 + \bar{Y}^2 + S_{O_2}^2 + \bar{O}_2^2) + \bar{O}_1^2(S_y^2 + S_{O_2}^2)}{\bar{O}_1^2(\bar{Y}^2 + \bar{O}_2^2 + 2\bar{O}_2\bar{Y})}$$

SRT-V: Diana and Perri [4] proposed a further refinement of the scrambling mechanism. According to their model, the sensitive variable y is transformed into z using

$$z = O_1[\chi_y O_2 + (1-\chi_y)y], \quad \chi_y \in [0,1). \tag{9}$$

The corresponding estimator for the population mean is:

$$\bar{Y}_{dp} = \frac{\bar{Z}_{dp} - \chi_y\bar{O}_2}{\chi_y + (1-\chi_y)\bar{O}_1}. \tag{10}$$

The associated expressions for correlation and coefficient of variation are

$$\rho_{zx} = \frac{\rho_{yx}\bar{O}_1 S_y(1-\chi_{y(opt)})}{\sqrt{a_0}}, \quad C_z^2 = \frac{a_0}{\bar{X}^2},$$

where

$$a_0 = \chi_{y(opt)}^2[S_{O_2}^2(S_{O_1}^2 + \bar{O}_1^2) + S_{O_1}^2\bar{O}_2^2] + (1-\chi_{y(opt)})^2[S_{O_1}^2(S_y^2 + \bar{Y}^2) + S_y^2\bar{O}_1^2]$$
$$+ 2\chi_{y(opt)}(1-\chi_{y(opt)})\bar{O}_2\bar{Y}S_{O_1}^2$$

and the optimal value of χ_y is given by

$$\chi_{y(opt)} = \left[\frac{S_{O_1}^2(S_y^2 + \bar{Y}^2 - \bar{O}_2\bar{Y})S_y^2\bar{O}_1^2}{S_{O_1}^2(S_{O_2}^2 + \bar{O}_2^2 + S_y^2 + \bar{Y}^2 - 2\bar{O}_2\bar{Y}) + S_{O_2}^2\bar{O}_1^2 + S_y^2\bar{O}_1^2}\right].$$

Remark 3.1. The scrambling variables O_1 and O_2 are assumed to satisfy the following conditions:

$$E(O_1) = \bar{O}_1, \quad E(O_2) = \bar{O}_2, \quad V(O_1) = \sigma_{O_1}^2, \quad V(O_2) = \sigma_{O_2}^2.$$

Remark 3.2. The notations $(\bar{Y}_j)$; $j \in \{ts, a, m, s, dp\}$ refer to the population mean of the sensitive variable y under the SRT-1, SRT-II, SRT-III, SRT-IV and SRT-V scrambling models, respectively.

Remark 3.3. To construct an appropriate estimator for the sensitive population mean $\bar{Y}$, it is first necessary to obtain a reliable estimator for the scrambled response mean $\bar{Z}$. This estimator is then substituted into eqs. (2), (4), (6), (8), and (10), corresponding to the different scrambling models. The following section is dedicated to the development of such an estimator for $\bar{Z}$.

3.1 Measurement Error in Scrambled Response Framework

We now extend the scrambled response model to account for measurement errors in both the observed scrambled response variable and the auxiliary variable. Let z_e and x_e denote the observed (error-prone) versions of the true scrambled response z and the auxiliary variable x, respectively. Under the classical additive measurement error model, their relationships are defined as:

$$z_{ei} = z_i + u_i,\ x_{ei} = x_i + v_i, \quad \text{for } i = 1, 2, \ldots, N,$$

where u_i and v_i are the measurement errors associated with z and x, respectively. These errors are assumed to be normally distributed with zero means and variances σ_u^2 and σ_v^2. Furthermore, the measurement errors in the study and auxiliary variables are allowed to be correlated.

4. Proposed Estimators in the Presence of Measurement Error

Horvitz-Thompson Estimator for the Scrambled Response Variable

In the context of the proposed sampling design, we aim to estimate the population mean of a sensitive variable under measurement error using the classical Horvitz-Thompson estimator (Horvitz and Thompson [7]). The adjusted form of this estimator, accounting for measurement error in the observed scrambled response, is given by

$$\hat{H}_h^{me} = \frac{1}{N} \sum_{i \in s_n} \gamma_i z_{ei}, \quad \text{where } \gamma_i = \frac{1}{\pi_i}, \tag{11}$$

where z_{ei} represents the observed (error-prone) scrambled response for unit i, π_i is the inclusion probability of unit i, and N is the total population size.

4.1 Calibration-Based Estimators under Measurement Error

Calibration is a widely used and powerful technique in survey sampling for improving parameter estimation, especially when auxiliary information is available. It enhances the precision and robustness of estimators by adjusting the survey weights so that they align with known population totals. The effectiveness of calibration heavily relies on the quality and relevance of auxiliary variables-particularly their accuracy, availability, and, most importantly, their correlation with the study variable.

When auxiliary variables are highly correlated with the variable of interest, calibration can substantially reduce the variance of estimators and mitigate non-sampling errors, such as measurement error. Deville and Sarndal [2] introduced a general calibration framework that uses a chi-square-type distance function to compute adjusted weights. Building upon their approach, calibration methods have since been extended to contexts involving measurement error and sensitive or coded responses.

To improve the performance of the traditional Horvitz-Thompson estimator, we apply a calibration approach wherein the original design weights γ_i are replaced by calibrated weights w_i, computed using known auxiliary information. In this context, we propose three calibration estimators for estimating the population mean in the presence of measurement error: a basic calibration estimator, a ratio-type calibration estimator, and an exponential-type calibration estimator,

$$H_C^{me} = \frac{1}{N}\sum_{i\in s_n} w_i z_{ei}, \tag{12}$$

$$H_R^{me} = \frac{1}{N}\sum_{i\in s_n} w_i z_{ei}\left(\frac{X_i}{x_{ei}}\right), \tag{13}$$

$$H_E^{me} = \frac{1}{N}\sum_{i\in s_n} w_i z_{ei}\exp\left(\frac{X_i - x_{ei}}{X_i + x_{ei}}\right). \tag{14}$$

To determine the calibrated weights w_i, we minimize the following chi-square-type distance function

$$\hat{\Theta}_1(w_i,\gamma_i) = \sum_{i=1}^{n}\frac{(w_i-\gamma_i)^2}{\gamma_i Q_i}, \tag{15}$$

subject to the calibration constraint:

$$\frac{1}{N}\sum_{i\in s_n} w_i x_{ei} = \bar{X}, \tag{16}$$

where Q_i is a known constant, and $\bar{X}$ is the population mean of the auxiliary variable.

The goal is to compute weights w_i that are as close as possible to the original design weights γ_i, while satisfying the constraint in eq. (16). This leads to a constrained optimization problem, which is solved by minimizing the following Lagrangian:

$$\Gamma_1^{me} = \hat{\Theta}_1(w_i,\gamma_i) - 2\lambda_1\left(\frac{1}{N}\sum_{i\in s_n} w_i x_{ei} - \bar{X}\right). \tag{17}$$

Differentiating Γ_1^{me} with respect to w_i and equating the derivative to zero yields

$$w_i = \gamma_i + \lambda_1 x_{ei}\alpha_i Q_i. \tag{18}$$

Solving for the Lagrange multiplier λ_1, we obtain

$$\lambda_1 = \frac{\bar{X} - \sum_{i\in s_n}\gamma_i x_{ei}}{\sum_{i\in s_n}\alpha_i Q_i x_{ei}^2}. \tag{19}$$

Substituting this value back into eq. (18), the calibrated weight becomes

$$w_i = \gamma_i + \gamma_i Q_i \frac{\left(\bar{X} - \sum_{i\in s_n}\gamma_i x_{ei}\right)x_{ei}}{\sum_{i\in s_n}\gamma_i Q_i x_{ei}^2}. \tag{20}$$

Finally, substituting w_i from eq. (20) into eqs. (12)-(14), we obtain the following explicit forms of the calibrated estimators under SRT in presence of measurement error as

$$\hat{H}_C^{me} = \frac{1}{N}\sum_{i\in s_n} z_{ei} + g_1\left[\bar{X} - \frac{1}{N}\sum_{i\in s_n}\gamma_i x_{ei}\right], \tag{21}$$

$$\hat{H}_R^{me} = \frac{1}{N}\sum_{i\in s_n} z_{ei}\left(\frac{X_i}{x_{ei}}\right) + g_2\left[\bar{X} - \frac{1}{N}\sum_{i\in s_n}\gamma_i x_{ei}\right], \tag{22}$$

$$\hat{H}_E^{me} = \frac{1}{N}\sum_{i\in s_n} z_{ei}\exp\left(\frac{X_i - x_{ei}}{X_i + x_{ei}}\right) + g_3\left[\bar{X} - \frac{1}{N}\sum_{i\in s_n}\gamma_i x_{ei}\right], \tag{23}$$

where the coefficients g_1, g_2, g_3 are given by

$$g_1 = \frac{\sum_{i\in s_n} x_{ei} z_{ei}\gamma_i Q_i}{\sum_{i\in s_n} x_{ei}^2\gamma_i Q_i}, \quad g_2 = \frac{\sum_{i\in s_n} x_{ei} z_{ei}\left(\frac{X_i}{x_{ei}}\right)\gamma_i Q_i}{\sum_{i\in s_n} x_{ei}^2\gamma_i Q_i}, \quad g_3 = \frac{\sum_{i\in s_n} x_{ei} z_{ei}\exp\left(\frac{X_i - x_{ei}}{X_i + x_{ei}}\right)\gamma_i Q_i}{\sum_{i\in s_n} x_{ei}^2\gamma_i Q_i}.$$

5. Analysis Under Simple Random Sampling Without Replacement (SRSWOR)

To examine the behaviour of the proposed calibration estimators within the framework of *Simple Random Sampling Without Replacement* (SRSWOR), we consider the standard inclusion probabilities associated with this sampling design. Under SRSWOR, the first-order and second-order inclusion probabilities for each unit are defined as

$$\pi_i = \frac{n}{N}, \quad \pi_{ij} = \frac{n(n-1)}{N(N-1)}, \quad i \neq j.$$

Assuming a constant value for the calibration parameter, i.e., $Q_i = 1$, the calibration estimators $\hat{H}_r^{me}$ for $r \in \{C, R, E\}$ under this design are denoted by $\hat{H}_r^{me*}$. These modified estimators take the following explicit forms:

$$\hat{H}_C^{me*} = \bar{z}_{en} + G_1(\bar{X} - \bar{x}_{en}), \qquad \text{where } G_1 = \frac{\sum_{i\in s_n} x_{ei} z_{ei}}{\sum_{i\in s_n} x_{ei}^2}, \tag{24}$$

$$\hat{H}_R^{me*} = \bar{z}_{en}\cdot\frac{\bar{X}}{\bar{x}_{en}} + G_2(\bar{X} - \bar{x}_{en}), \qquad \text{where } G_2 = \frac{\sum_{i\in s_n} x_{ei} z_{ei}\cdot\frac{X_i}{x_{ei}}}{\sum_{i\in s_n} x_{ei}^2}, \tag{25}$$

$$\hat{H}_E^{me*} = \bar{z}_{en}\cdot\exp\left(\frac{\bar{X} - \bar{x}_{en}}{\bar{X} + \bar{x}_{en}}\right) + G_3(\bar{X} - \bar{x}_{en}), \quad \text{where } G_3 = \frac{\sum_{i\in s_n} x_{ei} z_{ei}\cdot\exp\left(\frac{X_i - x_{ei}}{X_i + x_{ei}}\right)}{\sum_{i\in s_n} x_{ei}^2}, \tag{26}$$

where $\bar{z}_{en}$ and $\bar{x}_{en}$ denote the sample means of the observed scrambled response variable and the observed auxiliary variable under measurement error, respectively. These estimators incorporate both auxiliary information and calibration adjustments to enhance accuracy in the presence of measurement error.

6. Properties of the Proposed Estimators under Measurement Error

To evaluate the statistical properties of the proposed estimators $\hat{H}_r^{me*}$ for $r \in \{C, R, E\}$ under the presence of measurement error, we introduce the following notation:

$$K_0 = \frac{\bar{z}_{en}}{\bar{Z}} - 1, \; K_1 = \frac{\bar{x}_{en}}{\bar{X}} - 1 \text{ such that } E(K_i) = 0 \text{ for } i = 0, 1.$$

The corresponding second-order moments are given by

$$E(K_0^2) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_z^2 + S_u^2}{\bar{Z}^2},$$

$$E(K_1^2) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{S_x^2 + S_v^2}{\bar{X}^2},$$

$$E(K_0 K_1) = \left(\frac{1}{n} - \frac{1}{N}\right)\frac{\rho_{zx} S_z S_x + \rho_{uv} S_u S_v}{\bar{Z}\bar{X}}.$$

Using this notation, the Horvitz-Thompson estimator $\hat{H}_h^{me*}$ under SRSWOR can be expressed as

$$\hat{H}_h^{me*} = \bar{Z}(1+K_0),$$
$$\hat{H}_h^{me*} - \bar{Z} = \bar{Z}K_0.$$

Taking the square of the estimation error and expectation, we obtain the approximate variance (ignoring the finite population correction) given by

$$V(\hat{H}_h^{me*}) \approx \frac{1}{n}(S_u^2 + S_z^2). \tag{27}$$

Now, consider the calibrated estimator $\hat{H}_C^{me*}$ from eq. (24). Substituting in the notation gives

$$\hat{H}_C^{me*} = \bar{Z}(1+K_0) - G_1(\bar{X} - \bar{X}(1+K_1)), \tag{28}$$
$$\hat{H}_C^{me*} - \bar{Z} = \bar{Z}K_0 - G_1\bar{X}K_1. \tag{29}$$

Squaring both sides and applying first-order approximations yields

$$[\hat{H}_C^{me*} - \bar{Z}]^2 \approx [\bar{Z}K_0 - G_1\bar{X}K_1]^2. \tag{30}$$

Taking expectations on both sides of eq. (30), the variance of $\hat{H}_C^{me*}$ is

$$V[\hat{H}_C^{me*}] = \frac{1}{n}[S_z^2 + S_u^2 + G_1^2(S_x^2 + S_v^2) - 2BG_1(\rho_{zx}S_zS_x + \rho_{uv}S_uS_v)]. \tag{31}$$

Optimizing with respect to G_1, we obtain the minimum variance when

$$G_1^* = \frac{\rho_{zx}S_zS_x + \rho_{uv}S_uS_v}{S_x^2 + S_v^2}.$$

Substituting G_1^* into eq. (32), the minimum variance of the calibrated estimator becomes

$$V[\hat{H}_C^{me*}]_{opt} = \frac{1}{n}\delta_0, \tag{32}$$

where

$$\delta_0 = S_z^2 + S_u^2 + (G_1^*)^2(S_x^2 + S_v^2) - 2G_1^*(\rho_{zx}S_zS_x + \rho_{uv}S_uS_v).$$

Similarly, for the ratio and exponential-type estimators, the variances under optimal calibration weights G_2^* and G_3^* respectively are given as:

$$V[\hat{H}_R^{me*}]_{opt} = \frac{1}{n}[S_z^2 + S_u^2 + (G_2^*)^2(S_x^2 + S_v^2) + (S_x^2 + S_v^2) - 2G_2^*(\rho_{zx}S_zS_x + \rho_{uv}S_uS_v) - 2(\rho_{zx}S_zS_x + \rho_{uv}S_uS_v)], \tag{33}$$

where

$$G_2^* = \frac{2\rho_{uv}S_uS_v - \rho_{zx}S_zS_x}{S_x^2 + S_v^2}.$$

$$V[\hat{H}_E^{me*}]_{opt} = \frac{1}{n}\left[S_z^2 + S_u^2 + \frac{1}{4}(S_x^2 + S_v^2) + (G_3^*)^2(S_x^2 + S_v^2) + (S_x^2 + S_v^2) - 2G_3^*(\rho_{zx}S_zS_x + \rho_{uv}S_uS_v) - (\rho_{zx}S_zS_x + \rho_{uv}S_uS_v)\right], \tag{34}$$

with

$$G_3^* = \frac{2\rho_{uv}S_uS_v - \rho_{zx}S_zS_x - \frac{1}{2}(S_x^2 + S_v^2)}{S_x^2 + S_v^2}.$$

7. Estimators for the Sensitive Population Mean

To estimate the sensitive population mean, the population mean of the scrambled response variable $\bar{Z}$ in eqs. (2), (4), (6), (8), and (10) is replaced with the corresponding estimators $\hat{H}_h$, $\hat{H}_C$, $\hat{H}_R$, and $\hat{H}_E$. This substitution results in the respective estimators for the sensitive population mean, denoted by $\hat{\hat{Y}}_{jh}$, $\hat{\hat{Y}}_{jC}$, $\hat{\hat{Y}}_{jR}$, and $\hat{\hat{Y}}_{jE}$; $j \in \{ts, a, m, s, dp\}$.

The explicit forms of these estimators, along with their variances, are summarized in Table 1. The variances are derived based on the optimal conditions associated with each corresponding estimator of $\bar{Z}$.

Table 1. Estimators of the Sensitive Population Mean and their Variance Expressions

Model	Estimator	Variance Expression
SRT-I	$\hat{\hat{Y}}_{jr} = \frac{\hat{H}_r - \bar{O}_2}{\bar{O}_1}$	$V[\hat{\hat{Y}}_{jr}] = \frac{V(\hat{H}_r)}{\bar{O}_1^2}$
SRT-II	$\hat{\hat{Y}}_{jr} = \hat{H}_r - \bar{O}_2$	$V[\hat{\hat{Y}}_{jr}] = V(\hat{H}_r)$
SRT-III	$\hat{\hat{Y}}_{jr} = \frac{\hat{H}_r}{\bar{O}_1}$	$V[\hat{\hat{Y}}_{jr}] = \frac{V(\hat{H}_r)}{\bar{O}_1^2}$
SRT-IV	$\hat{\hat{Y}}_{jr} = \frac{\hat{H}_r - \bar{O}_1\bar{O}_2}{\bar{O}_1}$	$V[\hat{\hat{Y}}_{jr}] = \frac{V(\hat{H}_r)}{\bar{O}_1^2}$
SRT-V	$\hat{\hat{Y}}_{jr} = \frac{\hat{H}_r - \chi_y \bar{O}_2 \bar{O}_1}{(1-\chi_y)\bar{O}_1}$	$V[\hat{\hat{Y}}_{jr}] = \frac{V(\hat{H}_r)_{\text{opt.}}}{[(1-\chi_y)\bar{O}_1]^2}$

Note. $j \in \{ts, a, m, s, dp\}$ and $r \in \{h, C, R, E\}$

8. Simulation Study

To assess the performance of the proposed calibration estimators developed under the *Scrambled Response Technique* (SRT) in the presence of measurement error, and to compare them with existing estimators, a simulation study has been conducted.

For this purpose, a real-world population consisting of $N = 94$ districts from the southern states of India: Andhra Pradesh, Karnataka, Kerala, and Tamil Nadu has been considered (Population Source: URL: https://mohfw.gov.in).

The study involves the following variables:

- X_i: COVID-19 positivity rate in the ith district for the week of June 18th to June 24th, 2021.
- Y_i: COVID-19 positivity rate in the ith district for the week of June 26th to 2nd July, 2021.

The scrambling variables O_1 and O_2 are assumed to follow normal distributions: $O_1 \sim \mathcal{N}(1,1)$ and $O_2 \sim \mathcal{N}(1,2)$. Additionally, artificial measurement error terms u and v are generated independently from a normal distribution with mean 0 and variance 4, using MATLAB.

The known parameters of the population are

$$N = 94,\ \bar{Y} = 5.1756,\ \bar{X} = 5.1050,\ \rho_{YX} = 0.95, \quad \text{with sample sizes } n \in \{40, 55\}.$$

A Monte Carlo simulation with 10,000 independent replications was implemented in MATLAB to study the performance of the estimators. Specifically, the proposed calibration estimator $\hat{H}_C^{me}$ was compared with the Horvitz-Thompson estimator $\hat{H}_h^{me}$, the ratio-type calibration estimator $\hat{H}_R^{me}$, and the exponential-type calibration estimator $\hat{H}_E^{me}$ under the scrambling response models: SRT-I, SRT-II, SRT-III, SRT-IV and SRT-V, all in the presence of measurement error.

To evaluate and compare their efficiencies, *Percent Relative Efficiency* (PRE) values were computed as:

$$\text{PRE}_{j1} = \frac{V[\hat{\hat{Y}}_{jh}]}{V[\hat{\hat{Y}}_{jC}]} \times 100, \quad \text{PRE}_{j2} = \frac{V[\hat{\hat{Y}}_{jR}]}{V[\hat{\hat{Y}}_{jC}]} \times 100,$$

$$\text{PRE}_{j3} = \frac{V[\hat{\hat{Y}}_{jE}]}{V[\hat{\hat{Y}}_{jC}]} \times 100, \quad j = \begin{cases} ts, & \text{for SRT-I,} \\ a, & \text{for SRT-II,} \\ s, & \text{for SRT-III,} \\ m, & \text{for SRT-IV} \\ dp & \text{for SRT-V,} \end{cases}$$

where, for example,

$$V[\hat{\hat{Y}}_{tsh}] = \frac{1}{10{,}000} \sum_{i=1}^{10{,}000} [\hat{\hat{Y}}_{tshi} - \bar{Y}_i]^2.$$

Similar expressions are used to compute variances for other estimators under each scrambling model and for each estimator type $r_0 \in \{C, R, E\}$.

The simulation results for PRE_{jl}, where $l = 1, 2, 3$, are reported in Table 2.

Table 2. Simulation results for the percent relative efficiency of proposed calibrated estimators in the presence of measurement error under SRT models

j	Model	$n = 40$			$n = 55$		
		PRE_{j1}	PRE_{j2}	PRE_{j3}	PRE_{j1}	PRE_{j2}	PRE_{j3}
ts	SRT-I	107.2516	279.7016	33997	108.8433	479.2391	16969
a	SRT-II	143.9926	358.6410	26305	141.9766	499.0310	11152
m	SRT-III	106.7950	150.4416	11660	107.2640	124.1921	16274
s	SRT-IV	106.9398	4720.6340	17966	107.6941	2515.400	25016
	SRT-V:						
dp	χ_y	PRE_{j1}	PRE_{j2}	PRE_{j3}	PRE_{j1}	PRE_{j2}	PRE_{j3}
	0.1	109.6843	170.5627	18153	109.8137	168.6864	13881
	0.2	110.3145	213.4760	52090	110.3049	113.4960	42703
	0.3	110.1343	109.5357	10610	110.0753	105.0940	7752
	0.4	110.9766	110.3689	30414	110.3683	110.3683	21348
	0.5	110.0464	110.7634	14006	110.0233	110.3961	19023
	0.6	110.0652	120.4415	80593	110.0141	120.3036	64260
	0.7	110.0717	121.4301	52285	110.0069	121.3133	35913
	0.8	110.0087	122.5913	36109	110.0054	122.5946	25219
	0.9	110.0048	124.4503	26571	110.0049	124.4850	18180

In addition, to analyse the impact of measurement error, we compare the variances of the estimators obtained under SRT in presence of measurement error with those computed under an ideal scenario where no measurement error is present (i.e., setting $u = v = 0$). The corresponding relative efficiencies are computed as:

$$\mathrm{PRE}^*_{j1} = \frac{V[\hat{T}jc]}{V[\hat{\hat{Y}}_{jC}]} \times 100, \quad \mathrm{PRE}^*_{j2} = \frac{V[\hat{T}_{jr_1}]}{V[\hat{\hat{Y}}_{jR}]} \times 100,$$

$$\mathrm{PRE}^*_{j3} = \frac{V[\hat{T}_{je}]}{V[\hat{\hat{Y}}_{jE}]} \times 100, \quad j = \begin{cases} ts, & \text{for SRT-I,} \\ a, & \text{for SRT-II,} \\ s, & \text{for SRT-III,} \\ m, & \text{for SRT-IV} \\ dp & \text{for SRT-V,} \end{cases}$$

where, the estimators ($\hat{T}_{js}$, for $s \in \{c, r_1, e\}$, are derived from the respective $\hat{\hat{Y}}_j$ estimators by setting the measurement error components $u = v = 0$.

For instance,

$$V[\hat{T}_{tsc}] = \frac{1}{10,000} \sum_{i=1}^{10,000} [\hat{T}_{tsci} - \bar{Y}_i]^2.$$

Similar variance expressions are used for SRT-II, SRT-III, SRT-IV and SRT-V, and for all types of estimators.

The simulation outcomes for PRE^*_{jl}, where $l = 1, 2, 3$, are presented in Table 3.

Table 3. Simulation results for the percent relative efficiency of the proposed calibrated estimators without measurement error for $n = 55$

j	Model	PRE_{j1}	PRE_{j2}	PRE_{j3}
ts	SRT-I	0.1716	0.0294	0.0016983
a	SRT-II	0.5517	0.0672	0.0062104
m	SRT-III	0.0251	0.0022	0.0015876
s	SRT-IV	0.0262	0.0011	0.0010142
	SRT-V:			
	χ_y	PRE_{j1}	PRE_{j2}	PRE_{j3}
dp	0.1	0.8633	0.8322	0.01121
	0.2	1.0518	1.0105	0.01142
	0.3	1.2570	1.2036	0.01370
	0.4	1.5904	1.5154	0.011783
	0.5	2.0016	1.8977	0.01561
	0.6	2.7072	2.5468	0.01912
	0.7	3.6127	3.3613	0.011842
	0.8	5.3904	4.9402	0.012047
	0.9	9.1080	8.1624	0.012385

9. Discussion of Simulation Results

I. Table 2 presents the *Percent Relative Efficiency* (PRE) of the proposed calibrated estimators under various SRT models (SRT-I to SRT-V) in the presence of measurement error, for two different sample sizes: $n = 40$ and $n = 55$. Some important observations from Table 2 are:

- All PRE values across SRT models and both sample sizes are greater than 100, indicating that the proposed calibrated estimators are more efficient than the traditional estimators under measurement error conditions.
- In SRT-I and SRT-II, the exponential calibration estimator $\hat{H}_E^{me}$ shows significantly higher efficiency compared to the calibration estimator, with very large PRE_{j3} values.
- In SRT-V, where efficiency is evaluated across varying parameter of χ_y, a consistent trend is observed:

$$\text{PRE}_{dp1} < \text{PRE}_{dp2} < \text{PRE}_{dp3}.$$

  This clearly suggests that among all estimators, the exponential calibration estimator performs best, followed by the ratio calibration estimator.
- Comparing across models, the SRT-IV and SRT-V models, particularly at higher values of χ_y, yield higher PRE values, indicating that these models provide greater efficiency when used with exponential calibration.
- Larger sample size ($n = 55$) tends to produce slightly higher PREs, confirming the expectation that larger samples lead to more stable and efficient estimators.

These results confirm the applicability and effectiveness of the proposed calibrated estimators under SRT models with measurement error. Notably, the exponential calibration estimator under SRT-V exhibits the highest performance, making it a strong candidate for practical applications in sensitive variable estimation.

II. Table 3 evaluates the performance of the same proposed estimators in the absence of measurement error. The following trends can be noted:

- PRE values for SRT-I to SRT-IV are substantially less than 100, especially for exponential calibration, where PRE_{j3} values are extremely low. This indicates a considerable loss in efficiency when applying randomization and measurement error handling compared to using direct estimators.
- In SRT-V, as χ_y increases from 0.1 to 0.9, the PRE values for both PRE_{dp1} and PRE_{dp2} show an increasing trend, often exceeding 1. This implies that the calibration and ratio calibration estimators start to outperform their direct counterparts as the variation in χ_y increases.
- However, PRE_{j3} values for the exponential estimator remain very low throughout, even at higher χ_y levels, suggesting that the exponential calibration estimator is less efficient than its direct version in the absence of measurement error.

Overall, the results in Table 3 show that although the proposed estimators are useful under conditions of measurement error and sensitive data collection, their efficiency is lower than their direct counterparts when such issues are not present. Nevertheless, in sensitive surveys where direct questioning can lead to nonresponse or biased data, the use of randomized and measurement-error-robust estimators becomes essential despite the efficiency trade-off.

10. Concluding Remarks

The analysis of Tables 2 and 3 highlights the effectiveness of calibration based estimators for estimating the population mean of sensitive variables in the presence of measurement error. The proposed estimators-standard calibration, ratio calibration, and exponential calibration consistently demonstrate high percent relative efficiency, significantly outperforming the traditional Horvitz-Thompson estimator across various SRT models. Among these, the exponential calibration estimator under the SRT-V model emerges as the most robust, particularly at higher values of χ_y. However, a noticeable decline in precision is observed when comparing these estimators to their direct counterparts applied to clean data, indicating that while calibration techniques mitigate measurement error, they cannot fully recover lost efficiency. Despite this, the use of randomized response techniques remains crucial for obtaining truthful responses in sensitive surveys. Overall, the exponential calibration estimator within the scrambled response technique framework offers the best trade-off between statistical efficiency and respondent privacy, making it a practical and reliable choice for real-world applications.

Acknowledgement

Authors sincerely acknowledge the free access to data from https://www.mohfw.gov.in.

Competing Interests

The authors declare that they have no competing interests.

Authors' Contributions

All the authors contributed significantly in writing this article. The authors read and approved the final manuscript.

References

[1] W. E. Deming, *Some Theory of Sampling*, John Wiley and Sons, New York, xvii + 602 pages (1953).

[2] J.-C. Deville and C.-E. Särndal, Calibration estimators in survey sampling, *Journal of the American Statistical Association* **87**(418) (1992), 376 – 382, DOI: 10.1080/01621459.1992.10475217.

[3] G. Diana and P. F. Perri, A class of estimators for quantitative sensitive data, *Statistical Papers* **52**(3) (2011), 633 – 650, DOI: 10.1007/s00362-009-0273-1.

[4] G. Diana and P. F. Perri, New scrambled response models for estimating the mean of a sensitive quantitative character, *Journal of Applied Statistics* **37**(11) (2010), 1875 – 1890, DOI: 10.1080/02664760903186031.

[5] B. H. Eichhorn and L. S. Hayre, Scrambled randomized response method for obtaining sensitive quantitative data, *Journal of Statistical Planning and Inference* **7**(4) (1983), 307 – 316, DOI: 10.1016/0378-3758(83)90002-2.

[6] T. G. Gregoire and C. Salas, Ratio estimation with measurement error in the auxiliary variate, *Biometrics* **65**(2) (2009), 590 – 598, DOI: 10.1111/J.1541-0420.2008.01110.X.

[7] D. G. Horvitz and D. J. Thompson, A generalization of sampling without replacement from a finite universe, *Journal of the American Statistical Association* **47**(260) (1952), 663 – 685, DOI: 10.2307/2280784.

[8] P. C. Mahalanobis, Recent experiments in statistical sampling in the Indian Statistical Institute, *Sankhyā: The Indian Journal of Statistics* **20**(3/4) (1958), 329 – 398, URL: https://www.jstor.org/stable/25048402.

[9] P. F. Perri and G. Diana, Scrambled response models based on auxiliary variables, in: *Advances in Theoretical and Applied Statistics*, N. Torelli, F. Pesarin and A. Bar-Hen (editors), Springer-Verlag, Berlin, pp. 281 – 291 (2013), DOI: 10.1007/978-3-642-35588-2_26.

[10] K. H. Pollock and Y. Bek, A comparison of three randomized response models for quantitative data, *Journal of the American Statistical Association* **71**(336) (1976), 884 – 886, DOI: 10.1080/01621459.1976.10480963.

[11] K. Priyanka and P. Trisandhya, Calibration estimator in presence of measurement errors on successive occasions, *International Journal of Mathematics and Statistics* **24**(2) (2023), 1 – 14.

[12] K. Priyanka, P. Trisandhya and R. Mittal, Dealing sensitive characters on successive occasions through a general class of estimators using scrambled response techniques, *Metron* **76**(2) (2018), 203 – 230, DOI: 10.1007/s40300-017-0131-1.

[13] D. Raj, *Sampling Theory*, McGraw-Hill, New York, 302 pages (1968).

[14] A. Saha, A simple randomized response technique in complex surveys, *Metron* **65**(1) (2007), 59 – 66, URL: https://www.dss.uniroma1.it/RePec/mtn/articoli/2007-1-6.pdf.

[15] C.-E. Särndal, B. Swensson and J. Wretman, *Model Assisted Survey Sampling*, Springer-Verlag, New York, xv + 695 pages (1992), URL: https://link.springer.com/book/9780387406206.

[16] H. P. Singh and G. K. Vishwakarma, A general class of estimators in successive sampling, *Metron* **65**(2) (2020), 201 – 227, URL: https://www.dss.uniroma1.it/RePec/mtn/articoli/2007-2-4.pdf.

[17] P. V. Sukhatme, B. V. Sukhatme, S. Sukhatme and S. Ashok, *Sampling Theory of Surveys with Applications*, Indian Society of Agricultural Statistics, New Delhi, xiii + 530 pages (1984).

[18] S. L. Warner, Randomized response: A survey technique for eliminating evasive answer bias, *Journal of the American Statistical Association* **60**(309) (1965), 63 – 69, DOI: 10.2307/2283137.