AI IMPACT SUMMIT भारत 2026 INDIA
सर्वजन हिताय | सर्वजन सुखाय
WELFARE FOR ALL | HAPPINESS OF ALL

CASML 2025
Conference on AI for Scientific Machine Learning
Official Pre-Summit Event of the AI Impact Summit 2026

Some Calibration Estimators of the Mean of a Sensitive Variable under Measurement Error

Pidugu Trisandhya[1] Sat Gupta[2] and Frank Coolen[3]

[1]Department of Applied Sciences, Bharati Vidyapeeth's College of Engineering, New Delhi, India
[2]Department of Mathematics and Statistics, University of North Carolina at Greensboro, NC, USA
[3]Department of Mathematical Sciences, Durham University, Durham, UK

Abstract

This study examines the estimation of a sensitive population mean using calibration estimators under measurement error. It evaluates three randomized response techniques: Partial, Compulsory, and Optional. Theoretical properties are discussed, and a simulation based on real COVID-19 data is conducted. Results show that the Optional technique offers higher efficiency, highlighting its practical advantage in sensitive surveys.

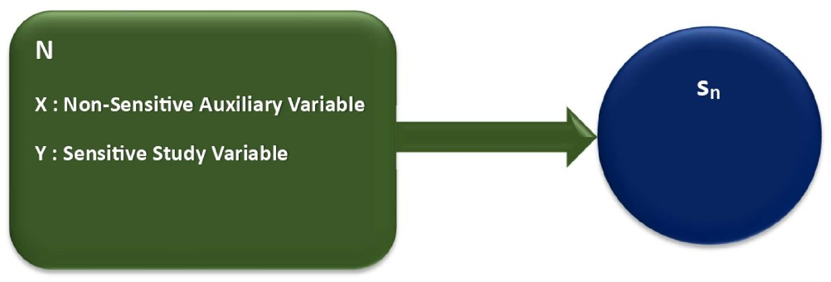
Introduction and Methodology

Introduction: Sensitive survey questions often yield biased responses. To address this, Warner (1965) introduced the Randomized Response Technique (RRT) for private yet reliable data collection, with later refinements by Horvitz & Thompson (1952), Greenberg *et al.* (1969), Franklin (1989), and Trisandhya *et al.* (2022) improving efficiency and participation.

Gupta *et al.* (2002) proposed the Optional RRT (ORRT), followed by the Partial RRT (PRRT) and other variants (Arcos *et al.* (2015); Priyanka *et al.* (2023)) for better handling of sensitive questions. Measurement error further impacts survey accuracy, with notable contributions from Mahalanobis (1946) and Priyanka *et al.* (2023).

This study applies CRRT, ORRT and PRRT models with measurement error, introducing basic calibration, ratio-type and exponential-type calibration estimators. A COVID-19 based simulation evaluates their performance under 3 scenarios: with measurement error, without measurement error, and without both randomization and measurement error.

Methodology: Consider a finite population $U = (U_1, U_2, \dots, U_N)$ of size N. Let Y denote the sensitive study variable and X a non-sensitive auxiliary variable with population means $\bar{Y}$ and $\bar{X}$, respectively. The objective is to estimate $\bar{Y}$ in the presence of measurement error. A sample s_n of size n is drawn using a sampling design d with inclusion probabilities $\pi_i = P(U_i \in s_n)$ and joint probabilities $\pi_{ij} = P(U_i, U_j \in s_n)$, where $\Delta_{ij} = \pi_{ij} - \pi_i\pi_j$. Within this framework, CRRT, ORRT, and PRRT are applied to manage sensitivity and assess the efficiency of different calibrated estimators under various randomization and measurement error conditions.



Sampling Design

Keywords and Sensitive Issues

Keywords:

- Compulsory/Scrambled response technique
- Sensitive variable
- Optional randomized response technique
- Calibration estimators
- Measurement Error.

Sensitive Issues:

1. Negligence of Government rules
2. Addiction of drugs
3. Criminal conviction
4. Acid attacks
5. Violence against women, etc.

Measurement Error: Assume the coded response Z and auxiliary variable X are observed with measurement errors. Let z_e and x_e denote the observed values. Using the classical additive error model: $z_{ei} = Z_i + u_i, x_{ei} = X_i + v_i, i = 1, 2, \dots, N$, where u_i and v_i are normally distributed with mean 0 and variances σ_u^2 and σ_v^2, possibly correlated.

Randomized Response Techniques

Let Z be the coded response for the sensitive variable Y, and S_1, S_2 be independent scrambling variables with means $\bar{S}_i$ and variances $\sigma^2_{S_i}$ $(i = 1, 2)$, independent of Y. Let P be the probability of a direct response (ORRT/PRRT). Following Zhang *et al.* (2021), respondents may report a direct or randomized response:

$$Z_j = \begin{cases} Y_j & \text{with probability } P \\ Y_j S_{1j} + S_{2j} & \text{with probability } 1 - P \end{cases}$$

$\bar{Z} = P\bar{Y} + (1-P)(\bar{Y}\bar{S}_1 + \bar{S}_2)$, giving the PRRT population mean:

$$\bar{Y} = \frac{\bar{Z} - (1-P)\bar{S}_2}{P + (1-P)\bar{S}_1}.$$

For $P = 0$ (CRRT): $Z_j = Y_j S_{1j} + S_{2j}$, $\quad \bar{Y} = \frac{\bar{Z} - \bar{S}_2}{\bar{S}_1}$.

Proposed ORRT:

$$Z_j = \begin{cases} Y_j & \text{with probability } P \\ \dfrac{Y_j S_{1j} + S_{1j} S_{2j} - \bar{S}_1 \bar{S}_2}{\bar{S}_1} & \text{with probability } 1 - P \end{cases}$$

$\bar{Y} = \bar{Z}$.

Here, $\bar{Y}$ can be estimated without knowing P.

Proposed Estimators in Presence of Measurement Error

HT-type Estimator in Presence of Measurement Error

$$\hat{\text{Ŧ}}_h^{me} = \frac{1}{N} \sum_{i \in s_n} \alpha_i z_{ei} \; ; \; \alpha_i = \frac{1}{\pi_i}. \tag{1}$$

Calibration Estimator in Presence of Measurement Error

$$\hat{\text{Ŧ}}_C^{me} = \frac{1}{N} \sum_{i \in s_n} w_i z_{ei}. \tag{2}$$

Ratio-type Calibration Estimator in Presence of Measurement Error

$$\hat{\text{Ŧ}}_R^{me} = \frac{1}{N} \sum_{i \in s_n} w_i z_{ei} \left(\frac{X_i}{x_{ei}} \right). \tag{3}$$

Exponential-type Calibration Estimator in Presence of Measurement Error

$$\hat{\text{Ŧ}}_E^{me} = \frac{1}{N} \sum_{i \in s_n} w_i z_{ei} \exp\left(\frac{X_i - x_{ei}}{X_i + x_{ei}} \right). \tag{4}$$

Estimators for Sensitive population mean under RRT

Table 1: Sensitive population mean estimator and their Variance in presence of measurement error. (where $j = C,\ R$ and E)

Model	**Estimators**	**Variance**
PRRT	$(\hat{\bar{Y}}_h)_{PRRT} = \frac{\hat{\text{Ŧ}}_h^{me*} - (1-P)\bar{S}_2}{P + (1-P)\bar{S}_1}$	$V[(\hat{\bar{Y}}_h)_{PRRT}] = \frac{V(\hat{\text{Ŧ}}_h^{me*})_{opt.}}{[P + (1-P)\bar{S}_1]^2}$
PRRT	$(\hat{\bar{Y}}_j)_{PRRT} = \frac{\hat{\text{Ŧ}}_j^{me*} - (1-P)\bar{S}_2}{P + (1-P)\bar{S}_1}$	$V[(\hat{\bar{Y}}_j)_{PRRT}] = \frac{V(\hat{\text{Ŧ}}_j^{me*})_{opt.}}{[P + (1-P)\bar{S}_1]^2}$
CRRT	$(\hat{\bar{Y}}_h)_{CRRT} = \frac{\hat{\text{Ŧ}}_h^{me*} - \bar{S}_2}{\bar{S}_1}$	$V[(\hat{\bar{Y}}_h)_{CRRT}] = \frac{V(\hat{\text{Ŧ}}_h^{me*})_{opt.}}{[\bar{S}_1]^2}$
CRRT	$(\hat{\bar{Y}}_j)_{CRRT} = \frac{\hat{\text{Ŧ}}_j^{me*} - \bar{S}_2}{\bar{S}_1}$	$V[(\hat{\bar{Y}}_j)_{CRRT}] = \frac{V(\hat{\text{Ŧ}}_j^{me*})_{opt.}}{[\bar{S}_1]^2}$
ORRT	$(\hat{\bar{Y}}_h)_{ORRT} = \hat{\text{Ŧ}}_h^{me*}$	$V[(\hat{\bar{Y}}_h)_{ORRT}] = V(\hat{\text{Ŧ}}_h^{me*})_{opt.}$
ORRT	$(\hat{\bar{Y}}_j)_{ORRT} = \hat{\text{Ŧ}}_j^{me*}$	$V[(\hat{\bar{Y}}_j)_{ORRT}] = V(\hat{\text{Ŧ}}_j^{me*})_{opt.}$

Simulation Study

A simulation study evaluated calibration estimators under three RRT models in the presence of measurement error using data from 94 districts in southern India (Andhra Pradesh, Karnataka, Kerala, and Tamil Nadu) were used [Source: https://www.mohfw.gov.in/state-data]. Weekly COVID-19 positivity rates for 18–24 June 2021 (X) and 21–27 June 2021 (Y) were analyzed with scrambling variables $S_1 \sim N(1, 1), S_2 \sim N(1, 2)$, and $u, v \sim N(0, 4)$. Using MATLAB, 10,000 replications were conducted to compare the proposed calibrated estimator with Horvitz-Thompson, ratio-type, and exponential-type estimators under three scenarios: with measurement error, without measurement error, and without both. Results, in terms of Percent Relative Efficiency (PRE), are shown in Tables 2 to 5.

Table 2: Simulation results for PRE of proposed calibrated estimators in presence of measurement error under PRRT model and ORRT model

P	PRRT						ORRT					
	$n = 30$			$n = 50$			$n = 30$			$n = 50$		
	PRE_{11}	PRE_{12}	PRE_{13}	PRE_{11}	PRE_{12}	PRE_{13}	PRE_{11}	PRE_{12}	PRE_{13}	PRE_{11}	PRE_{12}	PRE_{13}
0.1	100.66	174.12	253.95	126.39	224.05	312.53	197.12	212.71	414.00	198.04	255.00	11140
0.3	104.61	221.89	327.49	130.21	302.26	428.59	196.84	209.96	577.73	198.42	261.48	12949
0.5	100.00	240.67	402.10	100.89	310.12	467.81	197.10	218.70	609.62	198.53	276.86	17370
0.9	180.23	298.98	809.73	301.00	386.81	975.38	209.18	301.49	795.05	210.67	430.47	83663

Direct Method

Table 3: Simulation results for PRE of proposed calibrated estimators in absence of measurement error under PRRT model and ORRT model

P	PRRT						ORRT					
	$n = 30$			$n = 50$			$n = 30$			$n = 50$		
	PRE^*_{11}	PRE^*_{12}	PRE^*_{13}	PRE^*_{11}	PRE^*_{12}	PRE^*_{13}	PRE^*_{11}	PRE^*_{12}	PRE^*_{13}	PRE^*_{11}	PRE^*_{12}	PRE^*_{13}
0.1	95.03	51.12	20.00	90.46	37.22	20.04	101.11	39.89	20.14	102.97	36.80	20.09
0.3	115.84	39.56	20.04	137.63	36.82	20.04	101.79	39.27	20.11	103.11	36.34	20.07
0.5	117.27	38.11	20.03	148.56	37.03	20.03	102.03	38.64	20.08	104.48	35.72	20.06
0.7	101.14	35.84	20.02	123.95	35.75	20.02	102.79	37.45	20.05	103.96	34.81	20.03
0.9	34.87	31.73	20.00	52.39	31.81	20.00	72.49	33.53	20.01	75.66	32.26	20.00

Table 4: Simulation results for PRE of proposed calibrated estimators in presence of measurement error, in absence of measurement error, in absence of both measurement error and randomization under CRRT model

n	PRE_{11}	PRE_{12}	PRE_{13}	PRE^*_{11}	PRE^*_{12}	PRE^*_{13}	PRE^{**}_{11}	PRE^{**}_{12}	PRE^{**}_{13}
30	195.86	300.54	435.38	112.94	51.52	20.00	63.73	77.28	20.01
50	197.56	589.51	601.33	129.64	38.07	20.06	48.81	80.81	20.03

Table 5: Simulation results for PRE of proposed calibrated estimators in absence of measurement error and randomization with respect to PRRT and ORRT models for $n = 30$

P	PRRT			ORRT		
	PRE^{**}_{11}	PRE^{**}_{12}	PRE^{**}_{13}	PRE^{**}_{11}	PRE^{**}_{12}	PRE^{**}_{13}
0.1	30.06	20.06	10.00	64.17	30.29	20.16
0.3	30.10	20.00	10.01	56.98	21.27	20.11
0.5	30.20	20.01	10.01	93.27	13.60	20.06
0.7	30.54	20.04	10.02	14.91	26.91	20.02
0.9	33.30	20.16	10.05	28.85	21.42	20.00

Summarry Conclusion

Based on the analysis, estimating the population mean of a sensitive variable with measurement error is feasible. Calibration techniques improve efficiency under considered three models, with ORRT and PRRT performing better than CRRT. Among them, the proposed ORRT model is the most efficient when measurement error exists, though some loss in precision occurs compared to cases without measurement error or randomization. Thus, the ORRT model is recommended for estimating quantitative sensitive variables under measurement error.

Selected References

1 Warner, S. L. (1965): Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, **60**, 63-69.

2 Horvitz, D. G.; Thompson, D. J. (1952): A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, **47**, 663-685.

3 Greenberg, B. G.; Abul-Ela, A. L. A.; Simmons, W. R.; Horvitz, D. G. (1969): The unrelated question randomized response model: Theoretical framework. *Journal of the American Statistical Association*, **64**, 520-539.

4 Franklin, L. A. (1989): A comparison of estimators for randomized response sampling with continuous distributions from a dichotomous population. *Communications in Statistics - Theory and Methods*, **18**, 489-505.

5 Trisandhya, P.; Priyanka, K.; Kumar, A. (2022): Application of Item Sum Technique for estimating quantitative sensitive mean on successive moves using auxiliary information. *Communications in Statistics - Simulation and Computation*, **51**(7), 3868-3887.

6 Gupta, S.; Gupta, B.; Singh, S. (2002): Estimation of the sensitivity level of personal interview survey questions. *Journal of Statistical Planning and Inference*, **100**, 239-247.

7 Arcos, A.; Rueda, M.; Singh, S. A. (2015): A generalized approach to randomized response for quantitative variables. *Quality and Quantity*, **49**, 1239-1256.

8 Mahalanobis, P.C.(1946): Recent experiment in statistical sampling in the Indian statistical institute. *Journal of Royal Statistical Society, Series A.* **109** , 325-378.

9 Priyanka, K.; Trisandhya, P.; Kumar, A. (2023): Evaluating the effect of measurement error under randomized response techniques of the sensitive variable in successive sampling. *Proceedings of the National Academy of Sciences, Physical Sciences (NASA)*, **93**, 631-644. DOI: 10.1007/s40010-023-00836-w.

10 Zhang, Q.; Khalil, S.; Gupta, S. (2021): Mean estimation of sensitive variables under nonresponse and measurement errors using Optional RRT models. *Journal of Statistical Theory and Practice*, **15**, 1-15.