SCOPUA Journal of Applied Statistical Research (JASR)

Vol.2, Issue 3, September 2026
DOI: https://doi.org/10.64060/JASRv2i35

On the Analysis of Randomized Response Techniques under PPS Sampling Design

Muhammad Azeem [1*]

[1]Department of Statistics, University of Malakand, Khyber Pakhtunkhwa, Pakistan

* Corresponding Email: azeemstats@uom.edu.pk

Received: 21 July 2026 / Revised: 04 September 2026 / Accepted: 09 September 2026 / Published online: 15 September 2026

ABSTRACT

In recent years, exploration of randomized response techniques has been a topic of interest among survey statisticians. Researchers use randomized response methods to obtain data related to sensitive characteristics e.g. exam cheating, consumption of drugs, illegal income, tax payment, and violation of laws, etc. The available randomized response methods use the conventional equal probability sampling methods where all units share equal probabilities of selection in the sample. In many real-life problems, the probability of selection may vary from unit to unit, making it impracticable to use equal probability sampling designs in such situations. In such cases, the PPS (probability proportional to size) sampling scheme becomes an appropriate technique for selection of samples. This study develops the mathematical framework for the analysis of optional randomized response technique (ORRT) models in PPS sampling. Additionally, an improved randomized response model using forced responses has also been presented under PPS sampling. The improvement in the suggested model over the competitor models has been observed under PPS sampling.

Keywords*:*Efficiency; Estimator; PPS sampling; Privacy Protection; randomized response

1. Introduction

In many fields of life, researchers need to study the characteristics of sensitive variables. Survey statisticians use randomized response techniques for obtaining sensitive information from the respondents. Warner (1971) employed an additive randomized response approach to collect data on quantitative variables. Bar-Lev et al. (2004) suggested a randomized scrambling model, using a multiplicative-type variable to randomize the responses. The studies of Gjestvang and Singh (2009), Diana and Perri (2011), Narjis and Shabbir (2023), Saleem et al. (2023), Azeem (2023), Azeem and Ali (2023), Salemian et al. (2024), Gupta et al. (2024), and Pushadapu et al. (2025) explored different aspects of randomized response scrambling methods.

If the probability of selection varies from unit to unit, PPS sampling is used to draw a random sample from the population. The PPS sampling procedure was first introduced by Hansen and Hurvitz (1943) in the case of sampling with replacement. Later on, Horvitz and Thompson (1952) extended PPS sampling to sample selection without replacement. Narain (1951) and Sen (1953) also studied estimation of parameters under PPS sampling. Ahmad and Shabbir (2018) used extreme values to estimate population mean in PPS sampling. Sinha and Khanna (2022) and Ahmad et al. (2023a, 2023b) have also suggested mean estimators under PPS sampling. Recently, Khan et al. (2024) and Alomani et al. (2025) analyzed estimation methods in PPS sampling.

In many real-world surveys, the selection probability varies between units. For example, a statistician may be interested in selecting a random sample of students in such a manner that students with better academic grades have higher chances of selection. In such real-life situations, the appropriate method for sample selection is the PPS sampling design. If the researcher is interested in the analysis of sensitive quantitative variables such as corporate fraud, illegal earning, tax evasion, and cheating in examinations, randomized response methods are useful in gathering information from human respondents.

In the available literature, a few studies have attempted qualitative versions of randomized response models under PPS sampling for estimation of proportion. However, to our knowledge, there is rarity of quantitative Optional Randomized

Response techniques (ORRT) in PPS sampling. The rarity of randomized response models leaves a big research gap for the exploration and development of ORRT models under PPS sampling. The current study attempts to develop a mathematical framework for the analysis of ORRT models in PPS sampling. We extend a few of the existing quantitative models to the case of PPS sampling. Further, a new forced quantitative ORRT model has been developed, and its properties have been analyzed under PPS sampling.

Section 2 introduces some notations and also presents a review of the available models. Section 3 provides the description and properties of the proposed forced ORRT model. Section 4 presents the procedure for the assessment of privacy levels offered by ORRT models under PPS sampling. In Section 5, a real-world interview example using the suggested model has been provided. Section 6 presents a comparison of the proposed and competitor models. A detailed discussion of the findings has been presented in Section 7. Finally, Section 8 provides some recommendations for possible future work.

2. A Review of the Competitor Models

In this section, we present some of the available competitor models and the mean estimators under each model in PPS sampling. Before proceeding further, let us start with introducing some notations and assumptions used in this study.

Let the population and sample sizes be denoted by *N* and *n*, respectively. Let us denote the main variable and scrambling variable by *Y*, and *S*, respectively. Let μ_Y and μ_S denote the population mean of variables *Y* and *S*, respectively. Further, let σ_Y^2 and σ_S^2 denote the population variance of variables *Y* and *S*, respectively. Moreover, we also assume that *Y* and *S* are uncorrelated with each other.

Let

$$u_i = \frac{z_i}{NP_i}, \tag{1}$$

where P_i is the probability of the selection of *i*th population unit ($\sum_i P_i=1$), and z_i is the reported response under a given model. Moreover, let us denote:

$$D_S^2 = \frac{1}{N^2} Var\left(\frac{S_i}{P_i}\right),$$

and

$$D_y^2 = \frac{1}{N^2} Var\left(\frac{Y_i}{P_i}\right),$$

so that

$$Var\left(\frac{S_i}{P_i}\right) = N^2 D_S^2,$$

$$Var\left(\frac{Y_i}{P_i}\right) = N^2 D_y^2.$$

Also,

$$E\left(\frac{S_i}{P_i}\right) = N\mu_S = 0,$$

$$E\left(\frac{S_i^2}{P_i^2}\right) = Var\left(\frac{S_i}{P_i}\right) + \left[E\left(\frac{S_i}{P_i}\right)\right]^2 = N^2 D_S^2,$$

$$E\left(\frac{Y_i}{P_i}\right) = \sum_{i=1}^{N} \frac{Y_i}{P_i} P_i = \sum_{i=1}^{N} Y_i = N\mu_y,$$

and

$$E\left(\frac{Y_i^2}{P_i^2}\right) = Var\left(\frac{Y_i}{P_i}\right) + \left[E\left(\frac{Y_i}{P_i}\right)\right]^2 = N^2\left(D_y^2 + \mu_y^2\right).$$

Further, let *F* be a forced / fixed response, such that:

$$E\left(\frac{F}{P_i}\right)=NF\ ,$$

and

$$E\left(\frac{F^2}{P_i^2}\right)=N^2F^2\ .$$

The above notations and assumptions in the case of PPS sampling are consistent with the assumptions based on equal probability sampling. The independence between the main variable and the scrambling variables helps in privacy protection as the researcher cannot guess the true response from the values of the scrambling variable(s).

2.1 Warner's (1971) Additive Model

Warner's (1971) additive model is given as:

$$Z=Y_i+S_i\ . \tag{2}$$

The population mean may be estimated by the following unbiased estimator:

$$\hat{\mu}_W=\frac{1}{n}\sum_{i=1}^{n}u_i=\frac{1}{Nn}\sum_{i=1}^{n}\frac{z_i}{p_i}\ , \tag{3}$$

where z_i denotes the reported response using Eq. (2).

The variance of $\hat{\mu}_W$ is:

$$Var\left(\hat{\mu}_W\right)=Var\left(\frac{1}{Nn}\sum_{i=1}^{n}\frac{z_i}{p_i}\right),$$

or

$$Var\left(\hat{\mu}_W\right)=\frac{1}{N^2n^2}Var\left(\sum_{i=1}^{n}\frac{z_i}{p_i}\right)=\frac{1}{N^2n^2}\sum_{i=1}^{n}Var\left(\frac{z_i}{p_i}\right). \tag{4}$$

$Var\left(\frac{z_i}{p_i}\right)$ may be simplified as follows:

$$Var\left(\frac{z_i}{p_i}\right)=E\left(\frac{z_i}{p_i}\right)^2-\left[E\left(\frac{z_i}{p_i}\right)\right]^2. \tag{5}$$

Now,

$$E\left(\frac{z_i}{p_i}\right)=E\left(\frac{Y_i+S_i}{P_i}\right)=E\left(\frac{Y_i}{P_i}\right)+E\left(\frac{S_i}{P_i}\right),$$

or

$$E\left(\frac{z_i}{p_i}\right)=N\mu_y+0=N\mu_y\ . \tag{6}$$

Also,

$$E\left(\frac{z_i}{p_i}\right)^2=E\left(\frac{Y_i+S_i}{P_i}\right)^2=E\left(\frac{Y_i^2+S_i^2+2Y_iS_i}{P_i^2}\right),$$

or

$$E\left(\frac{z_i}{p_i}\right)^2=E\left(\frac{Y_i^2}{P_i^2}+\frac{S_i^2}{P_i^2}+2\frac{Y_iS_i}{P_i^2}\right),$$

or

$$E\left(\frac{z_i}{p_i}\right)^2=N^2\left(D_y^2+\mu_y^2\right)+N^2D_S^2\ . \tag{7}$$

Substituting Eq. (6) and Eq. (7) in Eq. (5) leads to:

$$Var\left(\frac{z_i}{p_i}\right) = N^2\left(D_y^2 + D_S^2\right). \tag{8}$$

Substituting Eq. (8) in Eq. (4) gives the sampling variance as:

$$Var\left(\hat{\mu}_W\right) = \frac{1}{n}\left[D_Y^2 + D_S^2\right]. \tag{9}$$

Under simple random sampling, the variance of $\hat{\mu}_W$ is:

$$Var_{SRS}\left(\hat{\mu}_W\right) = \frac{1}{n}\left[\sigma_Y^2 + \sigma_S^2\right].$$

From the above two equations, we see that the PPS version of the sampling variance can simply be obtained by substituting $\sigma_Y^2 = D_Y^2$ and $\sigma_S^2 = D_S^2$, which means that the notations and mathematical framework introduced in the start of Section 2 align with the conventional equal probability designs.

2.2 Gjestvang and Singh (2009) Technique

The ORRT model is described as follows:

$$Z = \begin{cases} Y_i - \beta S_i, & \text{with probability } \dfrac{\alpha}{\alpha+\beta}, \\ Y_i + \alpha S_i, & \text{with probability } \dfrac{\beta}{\alpha+\beta}, \end{cases} \tag{10}$$

where α and β are predefined constants.

The population mean may be estimated by the following unbiased estimator:

$$\hat{\mu}_{GS} = \frac{1}{n}\sum_{i=1}^{n} u_i = \frac{1}{Nn}\sum_{i=1}^{n}\frac{z_i}{p_i}, \tag{11}$$

where z_i denotes the reported response using Eq. (10). Using the method of derivations presented in Section 2.1, the variance of $\hat{\mu}_{GS}$ is:

$$Var\left(\hat{\mu}_{GS}\right) = \frac{1}{n}\left[D_Y^2 + \alpha\beta D_S^2\right]. \tag{12}$$

2.3 Narjis and Shabbir (2023) Technique

The scrambled response function may be expressed as:

$$Z = \begin{cases} Y_i - \beta S_i, & \text{with probability } \dfrac{\alpha}{\alpha+\beta+\gamma}, \\ Y_i + \alpha S_i, & \text{with probability } \dfrac{\beta}{\alpha+\beta+\gamma}, \\ Y_i, & \text{with probability } \dfrac{\gamma}{\alpha+\beta+\gamma}, \end{cases} \tag{13}$$

where α, β, and γ are predefined constants, such that $\alpha, \beta > 0$, and $0 < \gamma < 1$.

The population mean may be estimated by the following unbiased estimator:

$$\hat{\mu}_{NS} = \frac{1}{n}\sum_{i=1}^{n} u_i = \frac{1}{Nn}\sum_{i=1}^{n}\frac{z_i}{p_i}, \tag{14}$$

where z_i denotes the reported response using equation (13).

The variance of $\hat{\mu}_{NS}$ is:

$$Var\left(\hat{\mu}_{NS}\right) = \frac{1}{n}\left[D_Y^2 + \frac{\alpha\beta\left(\alpha+\beta\right)}{\alpha+\beta+\gamma}D_S^2\right]. \tag{15}$$

3. Proposed Model

Motivated by Narjis and Shabbir (2023), the following model is proposed:

$$Z=\begin{cases}(1+\gamma)Y_i-\beta S_i, & \text{with probability } \dfrac{\alpha}{\alpha+\beta+\gamma},\\ \left(1-\dfrac{\alpha\gamma}{\beta}\right)Y_i+\alpha S_i, & \text{with probability } \dfrac{\beta}{\alpha+\beta+\gamma},\\ F, & \text{with probability } \dfrac{\gamma}{\alpha+\beta+\gamma},\end{cases} \tag{16}$$

where F is a forced response. Under PPS sampling, an unbiased mean estimator is given as:

$$\hat{\mu}_P=\frac{\dfrac{1}{Nn}\sum_{i=1}^{n}\dfrac{Z_i}{P_i}-\dfrac{\gamma}{\alpha+\beta+\gamma}F}{\dfrac{\alpha+\beta}{\alpha+\beta+\gamma}}. \tag{17}$$

The properties of $\hat{\mu}_P$ are mathematically obtained in the following theorems.

Theorem 1: In PPS sampling, the estimator $\hat{\mu}_P$ is an unbiased estimator of μ_Y.

Proof: Applying expectation on Eq. (12) gives the equation:

$$E(\hat{\mu}_P)=E\left[\frac{\dfrac{1}{Nn}\sum_{i=1}^{n}\dfrac{Z_i}{P_i}-\dfrac{\gamma}{\alpha+\beta+\gamma}F}{\dfrac{\alpha+\beta}{\alpha+\beta+\gamma}}\right],$$

or

$$E(\hat{\mu}_P)=E_SE_R\left[\frac{\dfrac{1}{Nn}\sum_{i=1}^{n}\dfrac{Z_i}{P_i}-\dfrac{\gamma}{\alpha+\beta+\gamma}F}{\dfrac{\alpha+\beta}{\alpha+\beta+\gamma}}\right],$$

where E_S and E_R denote the expected values over samples and over the randomization device, respectively.

$$E(\hat{\mu}_P)=\frac{\dfrac{1}{Nn}\sum_{i=1}^{n}E_SE_R\left(\dfrac{Z_i}{P_i}\right)-\dfrac{\gamma}{\alpha+\beta+\gamma}F}{\dfrac{\alpha+\beta}{\alpha+\beta+\gamma}}. \tag{18}$$

$E_SE_R\left(\dfrac{Z_i}{P_i}\right)$ can be simplified as follows:

$$\begin{aligned}E_SE_R\left(\frac{Z_i}{P_i}\right)&=\frac{\alpha}{\alpha+\beta+\gamma}E_S\left[\frac{(1+\gamma)Y_i-\beta S_i}{P_i}\right]+\frac{\beta}{\alpha+\beta+\gamma}E_S\left[\frac{\left(1-\dfrac{\alpha\gamma}{\beta}\right)Y_i+\alpha S_i}{P_i}\right]+\frac{\gamma}{\alpha+\beta+\gamma}E_S\left(\frac{F}{P_i}\right),\\ &=\frac{\alpha}{\alpha+\beta+\gamma}\left[(1+\gamma)E_S\left(\frac{Y_i}{P_i}\right)-\beta E_S\left(\frac{S_i}{P_i}\right)\right]+\frac{\beta}{\alpha+\beta+\gamma}\left[\left(1-\frac{\alpha\gamma}{\beta}\right)E_S\left(\frac{Y_i}{P_i}\right)+\alpha E_S\left(\frac{S_i}{P_i}\right)\right]\\ &\quad+\frac{\gamma}{\alpha+\beta+\gamma}E_S\left(\frac{F}{P_i}\right),\\ &=\frac{\alpha}{\alpha+\beta+\gamma}\left[(1+\gamma)N\mu_Y-0\right]+\frac{\beta}{\alpha+\beta+\gamma}E_S\left[\left(1-\frac{\alpha\gamma}{\beta}\right)N\mu_Y+0\right]+\frac{\gamma}{\alpha+\beta+\gamma}NF,\end{aligned}$$

$$=\frac{\alpha+\beta}{\alpha+\beta+\gamma}N\mu_Y+\left[\alpha\gamma-\alpha\gamma\right]N\mu_Y+\frac{\gamma}{\alpha+\beta+\gamma}NF\,,$$

$$E_S E_R\left(\frac{Z_i}{P_i}\right)=\frac{\alpha+\beta}{\alpha+\beta+\gamma}N\mu_Y+\frac{\gamma}{\alpha+\beta+\gamma}NF\,. \tag{19}$$

Using Eq. (19) in Eq. (18) gives:

$$E\left(\hat{\mu}_P\right)=\frac{\frac{1}{Nn}\sum_{i=1}^{n}\left[\frac{\alpha+\beta}{\alpha+\beta+\gamma}N\mu_Y+\frac{\gamma}{\alpha+\beta+\gamma}NF\right]-\frac{\gamma}{\alpha+\beta+\gamma}F}{\frac{\alpha+\beta}{\alpha+\beta+\gamma}}\,,$$

or

$$E\left(\hat{\mu}_P\right)=\frac{\frac{\alpha+\beta}{\alpha+\beta+\gamma}\mu_Y+\frac{\gamma}{\alpha+\beta+\gamma}F-\frac{\gamma}{\alpha+\beta+\gamma}F}{\frac{\alpha+\beta}{\alpha+\beta+\gamma}}=\mu_Y\,. \tag{20}$$

Theorem 2: In PPS sampling, the variance of $\hat{\mu}_P$ is given as:

$$Var\left(\hat{\mu}_P\right)=\frac{1}{n\left(\alpha+\beta\right)^2}\left[\left(\alpha+\beta+\gamma\right)A\left(D_y^2+\mu_y^2\right)+\alpha\beta\left(\alpha+\beta\right)\left(\alpha+\beta+\gamma\right)D_S^2-\left\{\left(\alpha+\beta\right)\mu_Y+\gamma F\right\}^2\right],$$

where

$$A=\alpha\left(1+\gamma\right)^2+\beta\left(1-\frac{\alpha\gamma}{\beta}\right)^2.$$

Proof: Applying expectation on Eq. (17) gives:

$$Var\left(\hat{\mu}_P\right)=Var\left[\frac{\frac{1}{Nn}\sum_{i=1}^{n}\frac{Z_i}{P_i}-\frac{\gamma}{\alpha+\beta+\gamma}F}{\frac{\alpha+\beta}{\alpha+\beta+\gamma}}\right],$$

or

$$Var\left(\hat{\mu}_P\right)=\frac{\frac{1}{N^2n^2}\sum_{i=1}^{n}Var\left(\frac{Z_i}{P_i}\right)}{\left(\frac{\alpha+\beta}{\alpha+\beta+\gamma}\right)^2}=\frac{\frac{1}{N^2n}Var\left(\frac{Z_i}{P_i}\right)}{\left(\frac{\alpha+\beta}{\alpha+\beta+\gamma}\right)^2}. \tag{21}$$

Now,

$$Var\left(\frac{Z_i}{P_i}\right)=E\left(\frac{Z_i^2}{P_i^2}\right)-\left\{E\left(\frac{Z_i}{P_i}\right)\right\}^2. \tag{22}$$

$E\left(\frac{Z_i^2}{P_i^2}\right)$ can be simplified as follows:

$$E\left(\frac{Z_i^2}{P_i^2}\right)=E_S E_R\left(\frac{Z_i^2}{P_i^2}\right)=\frac{\alpha}{\alpha+\beta+\gamma}E_S\left[\frac{\left(1+\gamma\right)Y_i-\beta S_i}{P_i}\right]^2$$

$$+\frac{\beta}{\alpha+\beta+\gamma}E_S\left[\frac{\left(1-\frac{\alpha\gamma}{\beta}\right)Y_i+\alpha S_i}{P_i}\right]^2+\frac{\gamma}{\alpha+\beta+\gamma}E_S\left(\frac{F}{P_i}\right)^2,$$

$$=\frac{\alpha}{\alpha+\beta+\gamma}E_S\left[(1+\gamma)\left(\frac{Y_i}{P_i}\right)-\beta\left(\frac{S_i}{P_i}\right)\right]^2+\frac{\beta}{\alpha+\beta+\gamma}E_S\left[\left(1-\frac{\alpha\gamma}{\beta}\right)\left(\frac{Y_i}{P_i}\right)+\alpha\left(\frac{S_i}{P_i}\right)\right]^2$$

$$+\frac{\gamma}{\alpha+\beta+\gamma}E_S\left(\frac{F}{P_i}\right)^2,$$

$$=\frac{\alpha}{\alpha+\beta+\gamma}E_S\left[(1+\gamma)^2\left(\frac{Y_i}{P_i}\right)^2+\beta^2\left(\frac{S_i}{P_i}\right)^2-2\beta(1+\gamma)\left(\frac{Y_i}{P_i}\right)\left(\frac{S_i}{P_i}\right)\right]$$

$$+\frac{\beta}{\alpha+\beta+\gamma}E_S\left[\left(1-\frac{\alpha\gamma}{\beta}\right)^2\left(\frac{Y_i}{P_i}\right)^2+\alpha^2\left(\frac{S_i}{P_i}\right)^2+2\alpha\left(1-\frac{\alpha\gamma}{\beta}\right)\left(\frac{Y_i}{P_i}\right)\left(\frac{S_i}{P_i}\right)\right]+\frac{\gamma}{\alpha+\beta+\gamma}E_S\left(\frac{F}{P_i}\right)^2,$$

$$=\frac{\alpha}{\alpha+\beta+\gamma}\left[(1+\gamma)^2E_S\left(\frac{Y_i}{P_i}\right)^2+\beta^2E_S\left(\frac{S_i}{P_i}\right)^2-2\beta(1+\gamma)E_S\left(\frac{Y_i}{P_i}\right)\left(\frac{S_i}{P_i}\right)\right]$$

$$+\frac{\beta}{\alpha+\beta+\gamma}\left[\left(1-\frac{\alpha\gamma}{\beta}\right)^2E_S\left(\frac{Y_i}{P_i}\right)^2+\alpha^2E_S\left(\frac{S_i}{P_i}\right)^2+2\alpha\left(1-\frac{\alpha\gamma}{\beta}\right)E_S\left(\frac{Y_i}{P_i}\right)\left(\frac{S_i}{P_i}\right)\right]+\frac{\gamma}{\alpha+\beta+\gamma}E_S\left(\frac{F}{P_i}\right)^2,$$

$$=\frac{\alpha}{\alpha+\beta+\gamma}\left[(1+\gamma)^2N^2\left(D_y^2+\mu_y^2\right)+\beta^2N^2D_S^2-0\right]$$

$$+\frac{\beta}{\alpha+\beta+\gamma}\left[\left(1-\frac{\alpha\gamma}{\beta}\right)^2N^2\left(D_y^2+\mu_y^2\right)+\alpha^2N^2D_S^2+0\right]+\frac{\gamma}{\alpha+\beta+\gamma}N^2F^2,$$

$$E\left(\frac{Z_i^2}{P_i^2}\right)=\frac{1}{\alpha+\beta+\gamma}\left[\alpha(1+\gamma)^2+\beta\left(1-\frac{\alpha\gamma}{\beta}\right)^2\right]N^2\left(D_y^2+\mu_y^2\right)+\frac{\alpha\beta(\alpha+\beta)}{\alpha+\beta+\gamma}N^2D_S^2,$$

or

$$E\left(\frac{Z_i^2}{P_i^2}\right)=\frac{N^2}{\alpha+\beta+\gamma}\left[\left\{\alpha(1+\gamma)^2+\beta\left(1-\frac{\alpha\gamma}{\beta}\right)^2\right\}\left(D_y^2+\mu_y^2\right)+\alpha\beta(\alpha+\beta)D_S^2\right]. \tag{23}$$

Using Eq. (19) and Eq. (23) in Eq. (22) gives:

$$Var\left(\frac{Z_i}{P_i}\right)=\frac{N^2}{\alpha+\beta+\gamma}\left[\left\{\alpha(1+\gamma)^2+\beta\left(1-\frac{\alpha\gamma}{\beta}\right)^2\right\}\left(D_y^2+\mu_y^2\right)+\alpha\beta(\alpha+\beta)D_S^2\right]$$

$$-\frac{N^2}{(\alpha+\beta+\gamma)^2}\left[(\alpha+\beta)\mu_Y+\gamma F\right]^2.$$

Further simplification gives:

$$Var\left(\frac{Z_i}{P_i}\right)=\frac{N^2}{(\alpha+\beta+\gamma)^2}\left[(\alpha+\beta+\gamma)A\left(D_y^2+\mu_y^2\right)+\alpha\beta(\alpha+\beta)(\alpha+\beta+\gamma)D_S^2-\left\{(\alpha+\beta)\mu_Y+\gamma F\right\}^2\right]. \tag{24}$$

Using Eq. (24) in Eq. (20) gives the required result as:

$$Var(\hat{\mu}_P)=\frac{1}{n(\alpha+\beta)^2}\left[(\alpha+\beta+\gamma)A\left(D_y^2+\mu_y^2\right)+\alpha\beta(\alpha+\beta)(\alpha+\beta+\gamma)D_S^2-\left\{(\alpha+\beta)\mu_Y+\gamma F\right\}^2\right]. \tag{25}$$

4. Assessment of Respondents' Privacy

Every randomized response technique offers some degree of respondents' privacy protection which generally varies from model to model. To assess the level of privacy in a given technique, Yan et al. (2008) proposed a metric which is described as follows:

$$\Delta=E_SE_R(Z-Y)^2. \tag{26}$$

Many research studies have used the Yan et al. (2008) metric for assessment of privacy protection under equal probability sampling designs. In the case of PPS sampling design, we can simplify Eq. (26) as follows:

$$\Delta_{PPS}=\frac{\alpha}{\alpha+\beta+\gamma}E_S\left[(1+\gamma)Y_i-\beta S_i-Y_i\right]^2+\frac{\beta}{\alpha+\beta+\gamma}E_S\left[Y_i\left(1-\frac{\alpha\gamma}{\beta}\right)-Y_i+\alpha S_i\right]^2+\frac{\gamma}{\alpha+\beta+\gamma}E_S\left(F-Y_i\right)^2,$$

or

$$\Delta_{PPS}=\frac{\alpha}{\alpha+\beta+\gamma}\sum_{i=1}^{N}\left(\gamma Y_i-\beta S_i\right)^2 P_i+\frac{\beta}{\alpha+\beta+\gamma}\sum_{i=1}^{N}\left(\alpha S_i-\frac{\alpha\gamma}{\beta}Y_i\right)^2 P_i+\frac{\gamma}{\alpha+\beta+\gamma}\sum_{i=1}^{N}\left(F-Y_i\right)^2 P_i.$$

In the case of equal probability sampling designs, one may easily express the privacy measure Δ in the form of means / variances of different variables. However, under PPS sampling, we see that the sums of squares in the above equations are complex and cannot be put in the form of variances / means of variables. Alternatively, in order to align with the notations presented in Section 2, we can modify the privacy measure as:

$$\Delta_{PPS}=E_S E_R\left[\frac{Z_i-Y_i}{NP_i}\right]^2=\frac{1}{N^2}E_S E_R\left[\frac{Z_i-Y_i}{P_i}\right]^2. \tag{27}$$

The measure given in equation (27) aligns with equal probability sampling. If we put $P_i = 1/N$, equation (27) reduces to the privacy measure based on equal probability sampling.

Based on the metric given in Eq. (27) and using the notations given in Section 2, we can easily calculate the privacy protection level for any randomized response technique under PPS sampling. For instance, the privacy measure for the Warner's (1971), Gjestvang and Singh (2009), and Narjis and Shabbir (2023) techniques are obtained as:

$$\Delta_W=D_S^2, \tag{28}$$

$$\Delta_{GS}=\alpha\beta D_S^2, \tag{29}$$

and

$$\Delta_{NS}=\frac{\alpha\beta(\alpha+\beta)}{\alpha+\beta+\gamma}D_S^2. \tag{30}$$

To obtain the level of privacy for the proposed forced ORRT model in PPS sampling, we can proceed as follows:

$$\Delta_P=\frac{1}{N^2}E_S E_R\left[\frac{Z_i-Y_i}{P_i}\right]^2,$$

$$=\frac{1}{N^2}\left[\frac{\alpha}{\alpha+\beta+\gamma}E_S\left(\frac{(1+\gamma)Y_i-\beta S_i-Y_i}{P_i}\right)^2+\frac{\beta}{\alpha+\beta+\gamma}E_S\left(\frac{Y_i\left(1-\frac{\gamma\alpha}{\beta}\right)-Y_i+\alpha S_i}{P_i}\right)^2\right.$$

$$\left.+\frac{\gamma}{\alpha+\beta+\gamma}E_S\left(\frac{F-Y_i}{P_i}\right)^2\right],$$

$$=\frac{1}{N^2}\left[\frac{\alpha}{\alpha+\beta+\gamma}E_S\left(\frac{\gamma Y_i-\beta S_i}{P_i}\right)^2+\frac{\beta}{\alpha+\beta+\gamma}E_S\left(\frac{\alpha S_i-\frac{\alpha\gamma}{\beta}Y_i}{P_i}\right)^2+\frac{\gamma}{\alpha+\beta+\gamma}E_S\left(\frac{F-Y_i}{P_i}\right)^2\right],$$

$$=\frac{1}{N^2}\left[\frac{\alpha}{\alpha+\beta+\gamma}E_S\left(\gamma\frac{Y_i}{P_i}-\beta\frac{S_i}{P_i}\right)^2+\frac{\alpha\beta}{\alpha+\beta+\gamma}E_S\left(\frac{S_i}{P_i}-\frac{\gamma}{\beta}\frac{Y_i}{P_i}\right)^2+\frac{\gamma}{\alpha+\beta+\gamma}E_S\left(\frac{F}{P_i}-\frac{Y_i}{P_i}\right)^2\right],$$

$$\Delta_P=\frac{1}{N^2}\left[\frac{\alpha}{\alpha+\beta+\gamma}\left\{\gamma^2E_S\left(\frac{Y_i}{P_i}\right)^2+\beta^2E_S\left(\frac{S_i}{P_i}\right)^2-2\beta\gamma E_S\left(\frac{Y_i}{P_i}\frac{S_i}{P_i}\right)\right\}\right.$$

$$+\frac{\alpha\beta}{\alpha+\beta+\gamma}\left\{E_S\left(\frac{S_i}{P_i}\right)^2+\frac{\gamma^2}{\beta^2}E_S\left(\frac{Y_i}{P_i}\right)^2-2\frac{\gamma}{\beta}E_S\left(\frac{Y_i}{P_i}\frac{S_i}{P_i}\right)\right\}+\frac{\gamma}{\alpha+\beta+\gamma}\left\{E_S\left(\frac{F}{P_i}\right)^2+E_S\left(\frac{Y_i}{P_i}\right)^2-2E_S\left(\frac{Y_i}{P_i}\frac{F}{P_i}\right)\right\}\Bigg]\Bigg].$$

Further simplification yields:

$$\Delta_P=\frac{1}{N^2}\left[\frac{\alpha}{\alpha+\beta+\gamma}\left\{\gamma^2N^2\left(D_y^2+\mu_y^2\right)+\beta^2N^2D_S^2\right\}+\frac{\alpha\beta}{\alpha+\beta+\gamma}\left\{N^2D_S^2+\frac{\gamma^2}{\beta^2}N^2\left(D_y^2+\mu_y^2\right)\right\}\right.$$
$$\left.+\frac{\gamma}{\alpha+\beta+\gamma}\left\{N^2F^2+N^2\left(D_y^2+\mu_y^2\right)-2N^2F\mu_y\right\}\right],$$

or

$$\Delta_P=\frac{\alpha}{\alpha+\beta+\gamma}\left[\gamma^2\left(D_y^2+\mu_y^2\right)+\beta^2D_S^2\right]+\frac{\alpha\beta}{\alpha+\beta+\gamma}\left[D_S^2+\frac{\gamma^2}{\beta^2}\left(D_y^2+\mu_y^2\right)\right]$$
$$+\frac{\gamma}{\alpha+\beta+\gamma}\left[F^2+\left(D_y^2+\mu_y^2\right)-2F\mu_y\right],$$

or

$$\Delta_P=\frac{1}{\alpha+\beta+\gamma}\left[\alpha\gamma^2+\frac{\alpha\gamma^2}{\beta}+\gamma\right]\left(D_y^2+\mu_y^2\right)+\frac{\alpha\beta^2+\alpha\beta}{\alpha+\beta+\gamma}D_S^2+\frac{\gamma}{\alpha+\beta+\gamma}\left(F^2-2F\mu_y\right),$$

or

$$\Delta_P=\frac{1}{\alpha+\beta+\gamma}\left[B\left(D_y^2+\mu_y^2\right)+\alpha\beta\left(\beta+1\right)D_S^2+\gamma F\left(F-2\mu_y\right)\right], \tag{31}$$

where

$$B=\gamma\left(\alpha\gamma+\frac{\alpha\gamma}{\beta}+1\right).$$

It may be noted that we can easily obtain the privacy mean under equal probability designs by simply replacing D_S^2 and D_y^2 by σ_S^2 and σ_y^2, respectively, in Eq. (31). Likewise, we can also obtain the variance of the mean in the case of equal probability sampling designs by simply replacing D_S^2 and D_y^2 by σ_S^2 and σ_y^2, respectively, in Eq. (25). This means the notations introduced in Section 2 and the privacy measure presented in Eq. (27) align with their equal probability methods counterparts.

Gupta et al. (2018) suggested a unified metric for assessment of randomized response techniques. The unified measure may be expressed as:

$$\delta=\frac{Var\left(\hat{\mu}\right)}{\Delta}. \tag{32}$$

We can use the values of the sampling variance and the privacy measure in Eq. (32) to calculate the value of the unified measure for a given model.

5. Real-life Application

For the application to a real-life survey problem, a PPS sample consisting of 50 undergraduate students was selected and the respondents were interviewed regarding their parents' monthly income. Before data collection, the researcher explained the scrambling procedure to the respondents and ensured that the scrambling procedure was designed to protect their privacy. All of the respondents participated in our survey and reported their scrambled responses. The responses recorded are displayed in Table 1.

Table 1: Reported Responses

47,131	60,437	19,842	15,691	59,877	73,108	17,217	49,610	22,307	45,000
38,525	64,985	28,795	61,873	15,286	29,498	24,091	49,871	60,036	37,124
75,590	16,397	35,000	18,863	53,471	27,662	25,698	54,393	33,271	48,990
58,837	36,106	41,969	18,901	40,437	57,362	66,779	27,504	20,955	51,103
31,670	53,986	27,841	18,114	45,138	21,931	48,116	56,878	49,761	14,293

6. Comparative Analysis

We have calculated the variances of the mean and the corresponding values of the unified measure for the forced ORRT model and the competitor models, using different values of parameters. The results of the comparative performance of different models have been presented in tables. To show improvement, we have used hypothetical values of parameters. A discussion about the results has been given in next section.

Table 2: Values of variances and unified measure in PPS sampling using F = 100, $D_S^2 = 200$, $D_y^2 = 6$, $\mu_y = 50$, and $n = 500$

α	β	γ	Variance of the Mean			Unified Measure		
			$Var(\hat{\mu}_{GS})$	$Var(\hat{\mu}_{NS})$	$Var(\hat{\mu}_P)$	δ_{GS}	δ_{NS}	δ_P
1	2	0.01	0.81	0.81	0.76	0.0020	0.0020	0.0020
		0.05	0.81	0.80	0.58	0.0020	0.0020	0.0021
		0.10	0.81	0.79	0.34	0.0020	0.0020	0.0022
		0.15	0.81	0.77	0.11	0.0020	0.0020	0.0022
	3	0.01	1.21	1.21	1.18	0.0020	0.0020	0.0020
		0.05	1.21	1.20	1.04	0.0020	0.0020	0.0021
		0.10	1.21	1.18	0.87	0.0020	0.0020	0.0021
		0.15	1.21	1.17	0.71	0.0020	0.0020	0.0022
2	2	0.01	1.61	1.61	1.58	0.0020	0.0020	0.0020
		0.05	1.61	1.59	1.45	0.0020	0.0020	0.0021
		0.10	1.61	1.57	1.32	0.0020	0.0020	0.0021
		0.15	1.61	1.55	1.20	0.0020	0.0020	0.0022
	3	0.01	2.41	2.41	2.39	0.0020	0.0020	0.0020
		0.05	2.41	2.39	2.29	0.0020	0.0020	0.0021
		0.10	2.41	2.36	2.19	0.0020	0.0020	0.0021
		0.15	2.41	2.34	2.09	0.0020	0.0020	0.0021
3	2	0.01	2.41	2.41	2.39	0.0020	0.0020	0.0020
		0.05	2.41	2.39	2.30	0.0020	0.0020	0.0021
		0.10	2.41	2.36	2.23	0.0020	0.0020	0.0021
		0.15	2.41	2.34	2.19	0.0020	0.0020	0.0021
	3	0.01	3.61	3.61	3.59	0.0020	0.0020	0.0020
		0.05	3.61	3.58	3.53	0.0020	0.0020	0.0020
		0.10	3.61	3.55	3.47	0.0020	0.0020	0.0021
		0.15	3.61	3.52	3.43	0.0020	0.0020	0.0021

Table 3: Values of variances and unified measure in PPS sampling using F = 150, $D_S^2 = 500$, $D_y^2 = 30$, $\mu_y = 100$, and $n = 500$

α	β	γ	Variance of the Mean			Unified Measure		
			$Var(\hat{\mu}_{GS})$	$Var(\hat{\mu}_{NS})$	$Var(\hat{\mu}_P)$	δ_{GS}	δ_{NS}	δ_P
1	2	0.01	2.06	2.05	1.93	0.0021	0.0021	0.0021
		0.05	2.06	2.03	1.44	0.0021	0.0021	0.0022
		0.10	2.06	2.00	0.85	0.0021	0.0021	0.0023
		0.15	2.06	1.96	0.29	0.0021	0.0021	0.0026
	3	0.01	3.06	3.05	2.97	0.0020	0.0020	0.0021
		0.05	3.06	3.02	2.61	0.0020	0.0020	0.0021
		0.10	3.06	2.99	2.18	0.0020	0.0020	0.0022
		0.15	3.06	2.95	1.77	0.0020	0.0020	0.0022
2	2	0.01	4.06	4.05	3.97	0.0020	0.0020	0.0020
		0.05	4.06	4.01	3.65	0.0020	0.0020	0.0021
		0.10	4.06	3.96	3.34	0.0020	0.0020	0.0022
		0.15	4.06	3.92	3.12	0.0020	0.0020	0.0023
	3	0.01	6.06	6.05	5.99	0.0020	0.0020	0.0020
		0.05	6.06	6.00	5.75	0.0020	0.0020	0.0021
		0.10	6.06	5.94	5.50	0.0020	0.0020	0.0021
		0.15	6.06	5.89	5.31	0.0020	0.0020	0.0022
3	2	0.01	6.06	6.05	5.99	0.0020	0.0020	0.0020
		0.05	6.06	6.00	5.79	0.0020	0.0020	0.0021
		0.10	6.06	5.94	5.67	0.0020	0.0020	0.0021
		0.15	6.06	5.89	5.70	0.0020	0.0020	0.0023
	3	0.01	9.06	9.05	9.01	0.0020	0.0020	0.0020
		0.05	9.06	8.99	8.85	0.0020	0.0020	0.0021
		0.10	9.06	8.91	8.74	0.0020	0.0020	0.0021
		0.15	9.06	8.84	8.72	0.0020	0.0020	0.0022

7. Discussion and Conclusion

Survey researchers over the past several decades have been developing ORRT models using the conventional equal probability sampling schemes, where all units share equal probabilities of selection. In a majority of real-life sample surveys, the condition of equal selection probabilities may not be applicable, requiring the use of PPS sampling. Keeping in mind the rarity of ORRT models in PPS sampling in the available literature, we extended some of the existing equal probability sampling – based randomized response techniques to PPS sampling design. Additionally, a novel ORRT model using forced responses was also presented. We proved the unbiasedness of the mean estimator and obtained its variance under PPS sampling design. Moreover, the derivation of the privacy measure for randomized response techniques in the case of PPS sampling was discussed.

Using the option of a fixed / forced response, a novel ORRT model was introduced in Section 3 under PPS sampling scheme. The results of the sampling variance and the unified metric have been provided in Table 2 and Table 3, using different parameter values. Our results reveal that the proposed forced ORRT model produces smaller variances than the competitor models under PPS sampling. The tables also reveal that the values of the unified measure are almost equal for all three competitor models. Although in some cases, the proposed forced model produces slightly larger values of δ, the improved efficiency still makes the proposed model the best of all three competitor models. Therefore, the proposed model is recommended for use with real-world surveys where the selection probabilities differ between units.

8. Future Research Recommendations

For future researchers, the following recommendations are provided.

1. Future researchers may explore the efficiency of variance estimators under the proposed ORRT model in PPS sampling.
2. Researchers may also develop new ORRT models in the case of PPS sampling and analyze their properties using the expected values and notations used in our study.
3. Researchers may also investigate the effects of response and non-response errors on estimation of parameters under the proposed model in PPS sampling.

References

Ahmad, S., & Shabbir, J. (2018). Use of extreme values to estimate finite population mean under PPS sampling scheme. *Journal of Reliability and Statistical Studies*, 11(2), 99–112.

Ahmad, S., Zahid, E., Shabbir, J., Aamir, M., & Onyango, R. (2023a). Enhanced estimation of the population mean using two auxiliary variables under probability proportional to size sampling. *Mathematical Problems in Engineering*, Article ID 5564360. https://doi.org/10.1155/2023/5564360

Ahmad, S., Shabbir, J., Zahid, E., & Aamir, M. (2023b). Improved family of estimators for the population mean using supplementary variables under PPS sampling. *Science Progress*, 106(2): Article 00368504231180085. https://doi.org/10.1177/00368504231180085

Alomani, G., Kayid, M., & Ahmad, S. (2025). Application to radiation data sets by suggesting an improved mean estimator under probability proportional to size sampling. *Journal of Radiation Research and Applied Sciences*, 18, 101354. https://doi.org/10.1016/j.jrras.2025.101354

Azeem, M. (2023). Introducing a weighted measure of privacy and efficiency for comparison of quantitative randomized response models. *Pakistan Journal of Statistics*, 39(3), 377-385.

Azeem, M., & Ali, S. (2023). A neutral comparative analysis of additive, multiplicative, and mixed quantitative randomized response models. *PLOS ONE*, 18(4), e0284995. https://doi.org/10.1371/journal.pone.0284995

Bar-Lev, S. K., Bobovitch, E., & Boukai, B. (2004). A note on randomized response models for quantitative data. *Metrika*, 60(3), 255-260.

Diana, G., & Perri, P. F. (2011). A class of estimators of quantitative sensitive data. *Statistical Papers*, 52(3), 633-650. https://doi.org/10.1007/s00362-009-0273-1

Gjestvang, C. R., & Singh, S. (2009). An improved randomized response model: Estimation of mean. *Journal of Applied Statistics*, 36(12), 1361-1367. https://doi.org/10.1080/02664760802684151

Gupta, S., Mehta, S., Shabbir, J., & Khalil, S. (2018). A unified measure of respondent privacy and model efficiency in quantitative rrt models. *Journal of Statistical Theory and Practice*, 12(3), 506-511. https://doi.org/10.1080/15598608.2017.1415175

Gupta, S., Zhang, J., Khalil, S., & Sapra, P. (2024). Mitigating lack of trust in quantitative randomized response technique models. *Communications in Statistics – Simulation and Computation*, 53(6), 2624-2632. https://doi.org/10.1080/03610918.2022.2082477

Hansen, M. H., & Hurwitz, W. N. (1943). On the theory of sampling from a finite population. *Annals of Mathematical Statistics*, 14, 333-362.

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

Khan, S., Farooq, M., Ahmad, S., & Hussain, S. (2024). Improved estimator for estimation of population mean using predictive approach under PPS sampling. *VFAST Transactions on Mathematics*, 12(2), 1-16.

Narain, R. D. (1951). On sampling without replacement with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, 3, 169-174.

Narjis, G., & Shabbir, J. (2023). An efficient new scrambled response model for estimating sensitive population mean in successive sampling. *Communications in Statistics – Simulation and Computation*, 52(11), 5327-5344. https://doi.org/10.1080/03610918.2021.1986528

Pushadapu, K., Singh, S., & Sedory, S. A. (2025). An optimized optional randomized response technique. *International Statistical Review*, 93(1), 130-149. https://doi.org/10.1111/insr.12581

Saleem, I., Sanaullah, A., Al-Essa, L. A., & Bashir, S. (2023). Efficient estimation of population variance of a sensitive variable using a new scrambling response model. *Scientific Reports*, 13, Article ID 19913. https://doi.org/10.1038/s41598-023-45427-2

Salemian, H., Mahmoudi, E., Alamri, O. A., & Shabbir, J. (2024). Improving a novel quantitative randomized response method using auxiliary variable information. *Heliyon*, 10, e40367. http://dx.doi.org/10.1016/j.heliyon.2024.e40367

Sen, A. R. (1953). On the estimate of the variance in sampling with varying probabilities. *Journal of the Indian Society of Agricultural Statistics*, 5, 119-127.

Sinha, R. R., & Khanna, B. (2002). Estimation of population mean under probability proportional to size sampling with and without measurement errors. *Concurrency and Computation: Practice and Experience*, 34(18), Article e7023.

Warner, S. L. (1971). The linear randomized response model. *Journal of the American Statistical Association*, 66(336), 884-888. https://doi.org/10.1080/01621459.1971.10482364

Yan, Z., Wang, J., & Lai, J. (2008). An efficiency and protection degree-based comparison among the quantitative randomized response strategies. *Communications in Statistics – Theory and Methods*, 38(3), 400-408.

Declaration

Conflict of Study: The author has no conflict of interest to declare.

Funding Statement: The author received no funding for this study.

Availability of Data and Material: The data associated with our study have not been deposited into a publicly available repository. All relevant data is available within the article and its references.

Author(s) Bio / Authors' Note

Muhammad Azeem *holds PhD degree in Statistics with specialization in Survey Sampling. He is currently working as Assistant Professor in the Department of Statistics, University of Malakand, Pakistan. He has authored more than 50 peer-reviewed research publications in pure and applied Statistics. He has 12 years of post-PhD teaching experience at undergraduate and postgraduate level. He is also working as a referee for reputable impact factor journals.* Email*:* azeemstats@uom.edu.pk