Chapter 9:

A New Optimal Orthogonal Additive Randomized Response Model Based on Moments Ratios of Scrambling Variable

Tanveer A. Tarray

DOI: https://doi.org/10.21467/books.44.9

Additional information is available at the end of the chapter

Introduction

The Randomized response (RR) technique was first presented by Warner (1965) mainly to cut down the probability of (i) reduced response rate and (ii) inflated response bias experienced in direct or open survey relating to sensitive issues. Some recent involvement to randomized response sampling is given by Fox and Tracy (1986), Singh and Mathur (2004, 2005), Gjestvang and Singh (2006), Singh and Tarray (2013, 2014, 2015, 2016) and Tarray and Singh (2016, 2017, 2018) . We below give the description of the model due to Singh (2010):

Singh (2010) Additive Model

Let there be k scrambling variables denoted by S_j , j = 1,2,…,k whose mean θ_j (i.e. $E(S_j) = \theta_j$) and variance γ_j^2 (i.e. $V(S_j) = \gamma_j^2$) are known. In Singh's (2010) proposed optimal new orthogonal additive model named as (POONAM), each respondent selected in the sample is requested to rotate a spinner, as shown in Fig. 9.1, in which the proportion of the k shaded areas, say $P_1, P_2, \ldots P_k$ are orthogonal to the means of the k scrambling variables, say $\theta_1, \theta_2, \ldots, \theta_k$ such that:

$$\sum_{j=1}^{k} P_j \theta_j = 0 \tag{9.1}$$

and

$$\sum_{j=1}^{k} P_j = 1 \tag{9.2}$$

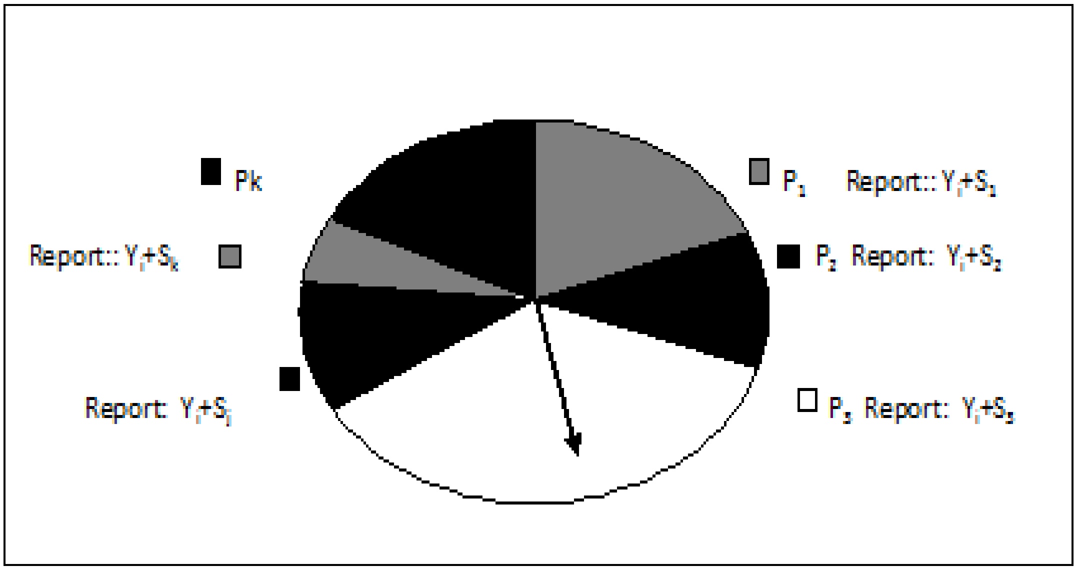

Figure 9.1: *Spinner for POONAM (Singh (2010))*

Now if the pointer stops in the j^{th} shaded area, then the i^{th} respondent with real value of the sensitive variable, say Y_i, is requested report the scrambled response Z_i as:

$$Z_i = Y_i + S_j \tag{9.3}$$

Assuming that the sample of size n is drawn from the population using simple random sampling with replacement (SRSWR). Singh (2010) suggested an unbiased estimator of the population mean μ_Y as

$$\hat{\mu}_Y = \frac{1}{n}\sum_{j=1}^{n} Z_j \tag{9.4}$$

The variance of $\hat{\mu}_Y$ is given by

$$V(\hat{\mu}_Y) = \frac{1}{n}\left[\sigma_y^2 + \sum_{j=1}^{k} P_j(\theta_j^2 + \gamma_j^2)\right] \tag{9.5}$$

The proposed procedure

It is to be noted that the mean θ_j and variance γ_j^2 of the j^{th} scrambling variable S_j (j=1,2,…,k) are known. Author has to propose a new additive model based on standardized scrambling variable $S_j^* = \left(\dfrac{S_j^2}{\theta_j(1+C_j^2)}\right)$, j = 1,2,…,k .

As demonstrated in Fig. 9.2, in which the proportion of the k shaded areas, say P_1, P_2, … P_k are orthogonal to the means of the k scrambling variables $(S_j^*, \forall\ j = 1,2,...,k)$, say $\theta_1, \theta_2, ..., \theta_k$ such that:

$$\sum_{j=1}^{k} P_j \theta_j = 0 \qquad (9.6)$$

and $$\sum_{j=1}^{k} P_j = 1 \qquad (9.7)$$

Now if the pointer stops in the j^{th} shaded area, then the i^{th} respondent with real value of the sensitive variable, say Y_i, is requested report the scrambled response Z_i^* as:

$$Z_i^* = Y_i + S_j^*$$

(9.8)

we prove the following theorems.

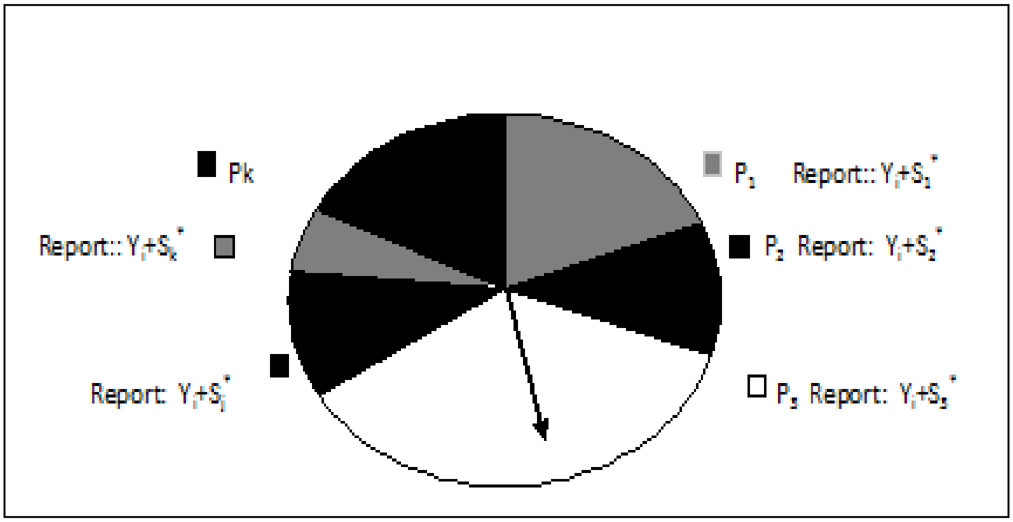

Figure 9.2: *Spinner for proposed procedure.*

Theorem 9.1

$$\hat{\mu}_{ST} = \frac{1}{n}\sum_{i=1}^{n} Z_i^* \tag{9.9}$$

Proof

Let E_1 and E_2 denote the expectations, then we have

$$E(\hat{\mu}_{ST}) = E_1 E_2\left[\frac{1}{n}\sum_{i=1}^{n} Z_i^*\right]$$

$$= E_1\left[\frac{1}{n}\sum_{i=1}^{n} E_2(Z_i^*)\right]$$

$$= E_1\left[\frac{1}{n}\sum_{i=1}^{n}\left\{Y_i\sum_{j=1}^{k}P_j + \sum_{j=1}^{k}P_j E_2(S_j^*)\right\}\right]$$

$$= E_1\left[\frac{1}{n}\sum_{j=1}^{n} Y_i\right] = \mu_Y \quad \text{, since } \sum_{j=1}^{k} P_j = 1 \text{ and } E_2(S_j^*) = \theta_j,$$

which proves the theorem.

Theorem 9.2

$$V(\hat{\mu}_{ST}) = \frac{1}{n}\left[\sigma_y^2 + \sum_{j=1}^{k}\frac{P_j A_j}{(1+C_j^2)^2}\right] \tag{9.10}$$

where

$$A_j = \left[\beta_2(S_j)C_j^4 + 4C_j^3 G_1(S_j) + 6C_j^2 + 1\right],$$

$\beta_2(S_j) = \dfrac{\mu_4(S_j)}{\gamma_j^4}$, $G_1(S_j) = \dfrac{\mu_3(S_j)}{\gamma_j^3}$ is the Fisher's measure of skewness, $\mu_3(S_j)$ and $\mu_4(S_j)$ are third and fourth central moments of the scrambling variable S_j.

Proof

$$V(\hat{\mu}_Y) = E_1 V_2(\hat{\mu}_Y) + V_1 E_2(\hat{\mu}_Y)$$

$$= E_1\left[V_2\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^*\right)\right] + V_1\left[E_2\left(\frac{1}{n}\sum_{i=1}^{n} Z_i^*\right)\right]$$

$$= E_1\left[\frac{1}{n^2}\sum_{i=1}^{n} V_2(Z_i^*)\right] + V_1\left[\frac{1}{n}\sum_{i=1}^{n} E_2(Z_i^*)\right]$$

$$= \frac{1}{n}\sum_{j=1}^{k}\frac{P_j A_j}{(1+C_j^2)^2} + \frac{\sigma_y^2}{n}$$

$$= \frac{1}{n}\left[\sigma_y^2 + \sum_{j=1}^{k}\frac{P_j A_j}{(1+C_j^2)^2}\right]$$

Note that:

$$V_2(Z_i^*) = E_2\left(Z_i^{*2}\right) - \left(E_2\left(Z_i^*\right)\right)^2$$

$$E_2(Z_i^{*2}) = E_2\left(Y_i + S_j^*\right)^2 = E_2\left[Y_i^2 + S_j^{*2} + 2Y_i S_j^*\right]$$

$$= Y_i^2\sum_{j=1}^{k}P_j + \sum_{j=1}^{k}P_j E_2(S_j^{*2}) + 2Y_i\sum_{j=1}^{k}P_j E_2(S_j^*)$$

$$= Y_i^2 + \sum_{j=1}^{k}P_j E_2(S_j^{*2}) + 2Y_i\sum_{j=1}^{k}P_j\theta_j$$

$$= Y_i^2 + \sum_{j=1}^{k}P_j E_2(S_j^{*2}),$$

Since $\sum_{j=1}^{k}P_j\theta_j = 0,$

and $E_2(S_j^{*2}) = E_2\left\{\frac{S_j^4}{\theta_j^2(1+C_j^2)^2}\right\}$

$$= \frac{1}{\theta_j^2(1+C_j^2)^2}E_2(S_j^4)$$

$$= \frac{1}{\theta_j^2(1+C_j^2)^2}E_2\left\{(S_j-\theta_j)^4 + 6\theta_j^2(S_j-\theta_j)^2 + 4\theta_j(S_j-\theta_j)^3 + 4\theta_j^2(S_j-\theta_j)\right\}$$

$$= \frac{1}{\theta_j^2(1+C_j^2)^2}\left\{\mu_4(S_j) + 4\theta_j\mu_3(S_j) + 6\theta_j^2\gamma_j^2 + \theta_j^4\right\}$$

$$= \mu_4(S_j) + 4\theta_j\mu_3(S_j) + 6\theta_j^2\gamma_j^2 + \theta_j^4$$

$$= \frac{\theta_j^4 A_j}{\theta_j^2(1+C_j^2)^2} = \frac{\theta_j^2 A_j}{(1+C_j^2)^2}$$

Thus $E_2(Z_i^{*2}) = Y_i^2 + \sum_{j=1}^{k} \frac{P_j A_j}{(1+C_j^2)^2}$

Therefore $V_2(Z_i^*) = Y_i^2 + \sum_{j=1}^{k} \frac{P_j A_j}{(1+C_j^2)^2} - Y_i^2$

$$= \sum_{j=1}^{k} \frac{P_j A_j}{(1+C_j^2)^2}$$

Efficiency Comparison

From (9.5) and (9.4), we have

$$V(\hat{\mu}_{ST}) < V(\hat{\mu}_Y) \text{ if}$$

i.e. if $\frac{1}{n}\left[\sigma_y^2 + \sum_{j=1}^{k} \frac{P_j A_j}{(1+C_j^2)^2}\right] < \frac{1}{n}\left[\sigma_y^2 + \sum_{j=1}^{k} P_j \theta_j^2 (1+C_j^2)\right]$

i.e if $\sum_{j=1}^{k} \frac{P_j A_j}{(1+C_j^2)^2} < \sum_{j=1}^{k} P_j \theta_j^2 (1+C_j^2)$

i.e if $\sum_{j=1}^{k} P_j \left[\frac{A_j}{(1+C_j^2)^2} - \theta_j^2(1+C_j^2)\right] < 0$

i.e. if $\frac{A_j}{(1+C_j^2)^2} < \theta_j^2(1+C_j^2) \qquad \forall\ j = 1,2,\ldots,k,$

i.e. if $A_j < \theta_j^2(1+C_j^2)^3 \qquad \forall\ j = 1,2,\ldots,k,$

i.e. if $\theta_j^2 > \frac{A_j^2}{(1+C_j^2)^3} \qquad \forall\ j = 1,2,\ldots,k. \qquad (9.11)$

ISBN: 978-81-936820-7-4

Book DOI: 10.21467/books.44

In case the scrambling variable S_j follows a normal distribution

(i.e. $S_j \sim N(\theta_j , \gamma_j^2)$, $j = 1,2,...,k$) , then A_j reduces to:

$$A_j^* = \left\{1 + 3C_j^2(2 + C_j^2)\right\} \quad (9.12)$$

Thus the condition (9.1) reduces to:

$$\theta_j^2 > \frac{(1 + 6C_j^2 + 3C_j^4)}{(1 + C_j^2)^3} \quad (9.13)$$

The condition (9.3) clearly indicates that $\left\{\theta_j^2 > \frac{(1 + 6C_j^2 + 3C_j^4)}{(1 + C_j^2)^3}, \forall\ j = 1,2,...,k\right\}$ then the proposed model is always better.

Further , suppose $S_j \sim N(\theta_j , \gamma_j^2)$, $\forall\ j = 1,2,...,k$, $\theta = 0$ and $\theta_j = 0$ $\forall\ j = 1,2,...,k$, then the variance expression in (9.5) and (9.4) respectively reduce to:

$$V(\hat{\mu}_Y) = \frac{1}{n}\left[\sigma_y^2 + \sum_{j=1}^{k} P_j \gamma_j^2\right] \quad (9.14)$$

and

$$V(\hat{\mu}_{ST}) = \frac{1}{n}\left[\sigma_y^2 + 3\right] \quad (9.15)$$

From (9.4) and (9.5) we have

$$V(\hat{\mu}_{ST}) - V(\hat{\mu}_Y) = \frac{1}{n}\left(\sum_{j=1}^{k} P_j \gamma_j^2 - 3\right)$$

$$= \frac{1}{n}\sum_{j=1}^{k} P_j\left(\gamma_j^2 - 3\right) \quad (9.16)$$

which is always positive if

$$\left(\gamma_j^2 - 3\right) > 0 \qquad \forall\ j = 1,2,...,k$$

i.e. if $\gamma_j^2 > 3 \qquad \forall\ j = 1,2,...,k$ (9.17)

Thus when $S_j \sim N(0 , \gamma_j^2)$, $\forall\ j = 1,2,...,k$, $\hat{\mu}_{ST}$ is more efficient as long as the condition (9.7) is satisfied.

In case S_j follows a normal distribution (i.e. $S_j \sim N(\theta_j, \gamma_j^2)$, $\forall\ j = 1,2,\ldots,k$) , PRE of $\hat{\mu}_{ST}$ with $\hat{\mu}_Y$ by using the formula:

$$PRE(\hat{\mu}_{ST}, \hat{\mu}_Y) = \frac{\left[\sigma_y^2 + \sum_{j=1}^{k} P_j \left\{(\theta_j^2 + \gamma_j^2)\right\}\right]}{\left[\sigma_y^2 + \sum_{j=1}^{k} \frac{P_j A_j^*}{(1+C_j^2)^2}\right]} \times 100 \tag{9.18}$$

where A_j^* is given in (9.2).

Suppose γ=40, γ_1= 30, γ_2= 40, γ_3 = 20, γ_4= 10, P_1=0.02, P_2=0.05, P_3=0.06, P_4=0.87 with k = 4 . $\sigma_y^2, \theta_1, \theta_2, \theta_3$ and θ_4 as listed in Table 9.1.

Table 9.1: $PRE(\hat{\mu}_{ST}, \hat{\mu}_Y)$

σ_Y^2	θ_1	θ_2	θ_3	θ_4	PRE
	300	200	100	-25.20	18523.16
	800	700	600	-100.00	242264.29
	1300	1200	1100	-174.70	732808.85
25	1800	1700	1600	-249.40	1490172.55
	300	200	100	-25.20	4130.07
	800	700	600	-100.00	53073.44
	1300	1200	1100	-174.70	160380.06
125	1800	1700	1600	-249.40	326053.37
	300	200	100	-25.20	2362.49
	800	700	600	-100.00	29839.47
	1300	1200	1100	-174.70	90081.79
225	1800	1700	1600	-249.40	183091.37
	300	200	100	-25.20	1672.71
325	800	700	600	-100.00	20772.56

	1300	1200	1100	-174.70	62648.32
	1800	1700	1600	-249.40	127301.32
	300	200	100	-25.20	1305.25
	800	700	600	-100.00	15942.52
	1300	1200	1100	-174.70	48034.22
425	1800	1700	1600	-249.40	97581.38
	300	200	100	-25.20	1076.99
	800	700	600	-100.00	12942.05
	1300	1200	1100	-174.70	38955.77
525	1800	1700	1600	-249.40	79119.00
	300	200	100	-25.20	921.41
	800	700	600	-100.00	10897.13
	1300	1200	1100	-174.70	32768.55
625	1800	1700	1600	-249.40	66536.36
	300	200	100	-25.20	808.58
	800	700	600	-100.00	9414.01
	1300	1200	1100	-174.70	28281.11
725	1800	1700	1600	-249.40	57410.48
	300	200	100	-25.20	723.01
	800	700	600	-100.00	8289.13
	1300	1200	1100	-174.70	24877.59
825	1800	1700	1600	-249.40	50488.93

From Table 9.1 $PRE(\hat{\mu}_{ST}, \hat{\mu}_Y)$ are greater than 100. It shows $\hat{\mu}_{ST}$ is more efficient than $\hat{\mu}_Y$ with substantial gain. Thus, the estimator $\hat{\mu}_{ST}$ over $\hat{\mu}_Y$ is recommended.

Table 9.2: *PRE of* $\hat{\mu}_{ST}$ *ove* $\hat{\mu}_Y$.

σ_y^2	25	125	225	325	425	525	625	725	825
PRE	835.71	260.94	190.35	162.80	148.13	139.02	132.80	128.30	124.88

The minimum values from 9.2 is observed as 124.88 and maximum 835.71 with a median of 148.13.

Table 9.2 PRE remains higher if the value of σ_y^2 is small. In that case the value of σ_y^2 will be around 0.5 to 5.0 (see Singh (2010), p. 67). It is observed that the PRE value decreases from 5985.71 to 2675.00 as the value of σ_y^2 increases from 0.5 to 5.0 .

Case k = 2 and the $PRE(\hat{\mu}_{ST}, \hat{\mu}_Y)$ for different parameters. Results are shown in Table 9.3.

Thus, the estimator $\hat{\mu}_{ST}$ over the estimator $\hat{\mu}_Y$ is recommended.

Table 9.3: *PRE of the estimator* $\hat{\mu}_{ST}$ *over the estimator* $\hat{\mu}_Y$ *with k =2.*

P_1	θ_1	θ_2	σ_Y^2	PRE
			25	1514232.14
			125	331316.41
			225	186046.05
			325	129355.18
			425	99155.37
			525	80394.89
			625	67609.08
			725	58335.85
0.2	1300	-325.0	825	51302.54
			25	219089.29
			125	48003.91
			225	26993.42
			325	18794.21
0.4	300	-200.0	425	14426.40

ISBN: 978-81-936820-7-4

Book DOI: 10.21467/books.44

			525	11713.07
			625	9863.85
			725	8522.66
			825	7505.43
			25	1528536.91
			125	334445.57
			225	187802.78
			325	130576.32
			425	100091.20
			525	81153.47
			625	68246.87
			725	58886.03
0.4	800	-533.3	825	51786.27
			25	1289517.86
			125	282160.16
			225	158449.56
			325	110172.26
			425	84454.44
			525	68478.22
			625	57589.97
			725	49692.99
0.8	300	-1200.0	825	43703.50

Conclusion

This paper elucidates amelioration over the Singh's (2010) randomized response model. We have advocated the optimal orthogonal additive randomized response model. The proposed model is found to be more resourceful both theoretically as well as numerically than the additive randomized response model studied by Singh (2010). Thus, the suggested RR procedure is therefore indorsed for its use in practice as an alternative to Singh's (2010) model.

Author's Detail

Tanveer A. Tarray

Department of Mathematical Sciences, Islamic University of Science and Technology, Kashmir, India -192122.

*Corresponding author email: tanveerstat@gmail.com

How to Cite this Chapter:

Tarray, Tanveer A. "A New Optimal Orthogonal Additive Randomized Response Model Based on Moments Ratios of Scrambling Variable." *Bayesian Analysis and Reliability Estimation of Generalized Probability Distributions*, edited by Afaq Ahmad, AIJR Publisher, 2019, pp. 95–107, ISBN: 978-81-936820-7-4, DOI: 10.21467/books.44.9.

References

Fox JA, Tracy PE (1986): *Randomized Response: A method of Sensitive Surveys.* Newbury Park, CA: SEGE Publications.

Gjestvang CR and Singh S (2006): A new randomized response model. Jour. Roy. Statist.Soc.68, 523-530.

Singh HP and Mathur N (2004): Estimation of population mean with known coefficient of variation under optional response model using scrambled response technique. Statist. Trans., 6, (7), 1079-1093.

Singh HP and Mathur N (2005): Estimation of population mean when coefficient of variation is known using scrambled response technique. Jour. Statist. Plan. Infer. 131, 135-144.

Singh HP and Tarray TA (2013): A modified survey technique for estimating the proportion and sensitivity in a dichotomous finite population. Inter.l Jour. Advanc. Scien. Techn. Res., 3(6), 459 – 472.

Singh HP and Tarray TA (2014): A dexterous randomized response model for estimating a rare sensitive attribute using Poisson distribution. Statist. Prob. Lett., 90, 42-45.

Singh, H.P., Tarray, T.A. (2015) : A revisit to the Singh, Horn, Singh and Mangat's randomization device for estimating a rare sensitive attribute using Poisson distribution. Model Assist. Stat. Appl. 10, 129–138.

Singh H.P. and Tarray T.A. (2017): Two - stage stratified partial randomized response strategies. Communications in Statistics-Theory & Methods, DOI: 10.1080/03610926.2013.804571.

Singh S (2010): Proposed optimal orthogonal new additive model (POONAM). Statistica, anno LXX (1), 73-81.

Tarray, T.A. and Singh, H.P. (2015). A randomized response model for estimating a rare sensitive attribute in stratified sampling using Poisson distribution. Model Assisted Statistics and Applications, 10, 345-360.

Tarray, T.A. and Singh, H.P. (2015). Some improved additive randomized response models utilizing higher order moments ratios of scrambling variable. Model Assisted Statistics and Applications, 10, 361-383.

Tarray, T.A. and Singh, H.P. (2016). New procedures of estimating proportion and sensitivity using randomized response in a dichotomous finite population. Journal of Model Applied Statistical Methods, 15(1), 635-669.

Tarray TA and Singh (2016): A stratified randomized response model for sensitive characteristics using the negative hyper geometric distribution. Comm. Statist. Theo. Metho., 45(4), 1014-1030, DOI: 10.1080/03610926.2013.853795.

Tarray, T.A. and Singh, H.P. (2017). A Survey Technique for Estimating the Proportion and Sensitivity in a Stratified Dichotomous Finite Population. Statistics and Applications, 15(1,2), 173-191.

Tarray T.A. and Singh H.P. (2017): A stratified randomized response model for sensitive

characteristics using the negative hyper geometric distribution. Communication in Statistics Theory- Methods. 46 (6), 2607-2629.

Tarray T.A. and Singh H.P. (2017): An optional randomized response model for estimating a rare sensitive attribute using Poisson distribution. Communication in Statistics Theory- Methods. 46 (6), 2638-2654.

Tarray T.A. and Singh H.P. (2018): A randomization device for estimating a rare sensitive attribute in stratified sampling using Poisson distribution. Afrika Matematika. DOI: 10.1007/s13370-018-0550-z.

Warner SL (1965): Randomized response: A survey technique for eliminating evasive answer bias. Jour. Amer. Statist. Assoc., 60, 63-69.