

A Field Test of a Mixture Binary Randomized Response Model to Estimate the Prevalence of a Sensitive Characteristic in the Finite Population

Samridhi Mehta¹, Neha Midha^{2*} and Priyanka Aggarwal²

¹ Department of Mathematics, Hindu College,
University of Delhi, Delhi 110007, India;
Email: smehta@hinducollege.du.ac.in

^{2*} Department of Statistics, Hindu College,
University of Delhi, Delhi 110007, India;

*Corresponding Author's Email: nmidha@hinducollege.du.ac.in

² Department of Statistics, Hindu College,
University of Delhi, Delhi 110007, India;
Email: priyankaaggarwal@hinducollege.ac.in

ABSTRACT

The objective of this paper is to validate a mixture binary randomized response model given in Lovig et al. (2021) through field work. The survey was conducted on students in the age group of 17-21 of a college affiliated to the University of Delhi, India. The students were asked two binary response questions, viz., 'Have you had a troubled childhood?' and 'Do you trust the model?'. The first question is the primary sensitive question and the latter question helps ascertain the proportion of respondents in the population who trust the randomized response technique model.

Keywords: Field Test, Binary Randomized Response Technique, Troubled Childhood, Confidential Survey, Face-to-Face Survey.

AMS Subject Classification: 62D05

1. INTRODUCTION

Childhood is often termed as a 'formative' period in a person's life, and rightfully so, due to the significant behavioural, emotional and cognitive capabilities that children develop at this age. They constantly observe and absorb the actions they see around them. Hence, childhood experiences and behaviour are leading indicators for a child's ability to deal with adulthood. While the upside of this tight correlation (that a happy child is likely have a strong foundation for adult life) is encouraging, the downside is that children with a troubled childhood can have negative, sometimes irreversible, long-term repercussions.

Instances of child abuse and neglect leave a lasting impression on a child. Child abuse could be physical, psychological, emotional, verbal or sexual. Child neglect includes physical neglect (failure to provide basic necessities of food, clothing and shelter), emotional neglect (not providing love, comfort,

or affection), and medical or educational neglect (not providing access to needed medical care or education) or supervisory neglect (failure to appropriately supervise). The impact of childhood experiences extends to all aspects of development including, but not limited to, brain development, social skills, ability to handle relationships, and mental health. In fact, research has shown that child trauma survivors may experience learning problems, including poorer academic performances with more suspensions and expulsions, low self-esteem and long-term health problems (e.g., psychiatric issues, alcoholism, depression, diabetes and heart disease) arising out of stress and prior life experiences. Further, the more adverse experiences in childhood may lead to developmental delays and other problems. Tomlinson et al. (2022), Downey and Crummy (2022), Lorenc et al. (2020) conducted research on the impact of adverse childhood experiences on the well-being of the individual.

It is believed that it is important to know the proportion of children having troubled childhood. This could facilitate policymakers in fostering the kinds of environments and experiences that provide children with a safety net they can turn to in the event of a problem. Efforts can also be made to prevent problems and remediate early difficulties so that the developmental process of children is not hampered. Thus, the selected sensitive question for our survey is 'Have you had a troubled childhood?'. This question addresses a number of issues such as sexual abuse, bullying, family issues, domestic violence, financial insecurities.

The attempt to gather information regarding the extent of prevalence of a sensitive attribute in a population is a challenging endeavour. The direct questioning of the respondent about possessing the sensitive characteristic could lead to biased responses that project the respondent in a socially desirable manner. To overcome this problem, the technique of randomized response technique (RRT) is adopted. The technique involves asking the respondents to use a random device, such as a deck of cards or a coin flip, whose outcome is unobserved by the interviewer. By introducing such a kind of random noise, the method conceals individual responses and protects respondent privacy, and simulated data is frequently used to assess the method's efficiency. In the literature, several randomized response technique (RRT) models exist. In real world situations, the RRT models can be easily used to estimate the prevalence of a sensitive characteristic in a population and this can be done through an actual survey. Thus, we aim to conduct a survey using our chosen RRT model to estimate a socially relevant sensitive characteristic 'Have you had a troubled childhood?'. In order to gather responses through a RRT model, we use the binary mixture RRT model of Lovig et al. (2021). This model combines the essence of both Warner (1965) model and Greenberg et al. (1969) model. Although, not many tread the path of validating the RRT model through a survey due to constraints of time, money and manpower, however, there exists research studies validating RRT models by using them to study the sensitive behaviour. Solomon et al. (2007) used RRT model to estimate illegal resource use at a Ugandan park. Ostopczuk et al. (2009) showcased a field work validation of RRT model, aimed at examining potential associations between education and attitude towards foreigners. Gill et al. (2013) employed binary and quantitative response models in surveys to study risky sexual behaviour. Chhabra et al. (2016) estimated the prevalence of sexual abuse by an acquaintance using optional unrelated question RRT model of Sihtm et al. (2016).

Through this work, an attempt is being made to validate the selected RRT model via a field test. The results obtained will be compared with the results obtained using a face-to-face interviewing technique and also with another technique that ensures response confidentiality. In Section 2 of this paper, the models used in this study are described. Section 3 outlines the survey methodology and its results which is followed by a discussion on the findings and references.

2. RANDOMISED RESPONSE TECHNIQUE MODELS

Let π_x be the proportion of the sensitive trait which is to be estimated for a population. Let $\widehat{\pi}_x$ denote its estimator. All the three RRT models below showcase a technique to estimate it.

2.1 Warner (1965) Model

Using a randomization mechanism, with predetermined probabilities, the respondents in this model are presented with a binary sensitive question or an indirect version of the same question. It is not known to the interviewer whether the respondent is answering the direct question or the indirect question.

Using the Warner (1965) model, $\widehat{\pi}_x = \frac{\widehat{P}_{yW} - (1-p)}{2p-1}$, $p \neq \frac{1}{2}$ where p and $1-p$ denotes the probability that the respondent is in the direct question group, indirect question group, respectively. Further, assume that $\widehat{P}_{yW}$ represents the estimate of the probability of 'yes' responses obtained using this model.

2.2 Greenberg et al. (1969) Model

Respondents are asked with predetermined probabilities to select randomly between the sensitive question or an entirely unrelated innocuous question in this binary model. The interviewer is unaware of the question the interviewee is replying to because the questions are asked using a randomization mechanism. In the Greenberg et al. (1969) model, let p and $1-p$ denote the probability that the respondent is in the direct question group, unrelated innocuous question group respectively. Assume π_y denote the known proportion of the unrelated characteristic and $\widehat{P}_{yG}$ represents the estimate of the probability of 'yes' response obtained using this model. Using the Greenberg et al. (1969) model, $\widehat{\pi}_x = \frac{\widehat{P}_{yG} - (1-p)\pi_y}{p}$

2.3 Binary Mixture Lovig et al. (2021) Model

Lovig et al. (2021) Model introduced a binary mixture RRT model that incorporate aspects of the above two binary models as shown in Figure 1. With predetermined probabilities, a respondent selects at random from the direct question, the indirect question, and the unrelated innocuous question. Additionally, Lovig et al. (2021) looked into the problem of untruthful responses in RRT models.

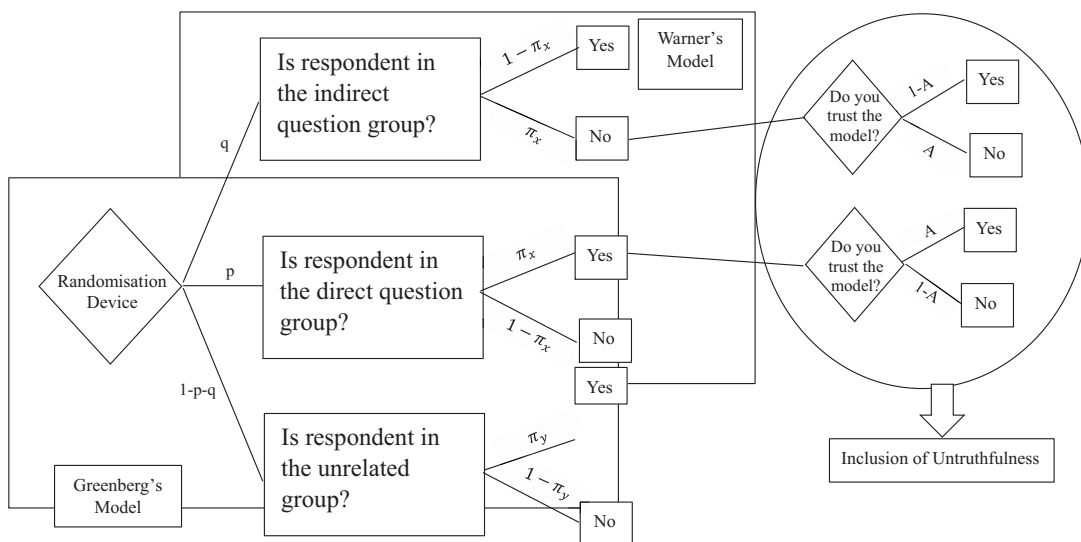


Figure 1. Lovig et al. (2021) model

Let p, q and $1 - p - q$ denote the probability that the respondent was in the direct question group, indirect question group and unrelated question group respectively. Assume that π_y denotes the known proportion of the unrelated innocuous characteristic. Let A and $1 - A$ respectively denote the proportion of people who would respond truthfully and who wouldn't respond truthfully to a sensitive question. Let n denote the sample size drawn with replacement. Let P_y denote the probability of a 'yes' response.

A two-question approach was used in the binary mixture Lovig et al. (2021) model to address the issue of untruthfulness. Each respondent is asked to respond to two questions through two randomization devices. An estimate of A is obtained using Question 1 through a randomization device which is based on Greenberg et al. (1969) Model wherein p_0 represents the proportion of the direct question used to estimate truthfulness and π_{y0} represent the proportion of people who would respond 'yes' to the unrelated question. It is to be noted that here the sensitive characteristic is 'trust the model'.

Thus, $P_{yG} = p_0A + (1 - p_0)\pi_{y0}$ yielding the unbiased estimator $\hat{A}$ of A ,

$$\hat{A} = \frac{\hat{P}_{y0} - (1 - p_0)\pi_{y0}}{p_0} \quad (1)$$

The second randomization device helps estimate the prevalence of the primary sensitive characteristic π_x .

$$P_y = \pi_x A(p - q) + q + (1 - p - q)\pi_y$$

$$\text{Using } \hat{P}_y \text{ which is an unbiased estimator of } P_y, \text{ we have } \hat{\pi}_x = \frac{\hat{P}_y - q - (1 - p - q)\pi_y}{\hat{A}(p - q)}, p \neq q \quad (2)$$

In order to find approximate value of $\hat{\pi}_x$ given in (2), Lovig et al. (2021) used first order Taylor's approximation and obtained an unbiased estimator for π_x given by

$$\hat{\pi}_x = \frac{P_y - q - (1 - p - q)\pi_y}{A(p - q)} - (\hat{A} - A) \left[\frac{P_y - q - (1 - p - q)\pi_y}{A^2(p - q)} \right] + \frac{(P_y - P_y)}{A(p - q)}, p \neq q \quad (3)$$

3. SURVEY METHODOLOGY AND ITS RESULTS

Field survey was conducted on 484 students of a college affiliated to the University of Delhi, Delhi, India. The participation in the survey was entirely voluntary and all the respondents were asked the main sensitive question through three methods: Direct face-to-face questioning, anonymous responding through ballot and binary mixture RRT model of Lovig et al. (2021).

Method 1 (Face-to-Face Method): In a face-to-face interview, the interviewer directly asks the respondents the sensitive question and the response given by the respondent is noted. 78 students out of 484 responded positively to the sensitive question and 406 students expressed that they have never experienced troubled childhood. Thus, 16.12% of respondents acknowledged experiencing a troubled childhood.

Method 2 (Confidential Method): In a confidential method, the respondent is asked to respond to the question anonymously. 129 'yes' responses and 355 'no' responses were received. So, 26.65% of respondents responded in agreement with having a troubled childhood. This increase in the estimate is well-expected as the respondents are not apprehensive of being judged by the interviewer when responding through a confidential method.

Method 3 (RRT Model): For the RRT model to yield meaningful results, it is important that the respondents comprehend the model well and they trust that when they respond through the model then their privacy will be maintained. In order to gain the respondent's trust, a brief session was conducted with the respondents to explain the concept of RRT and RRT models being used in the survey. A handout was also given to them that contained detailed information about the survey. They were assured that the privacy of their response will be maintained as their true responses regarding the possession of the sensitive characteristic are impossible to track down.

The survey was aimed to ascertain the prevalence of the sensitive attribute i.e. having a troubled childhood and also to estimate the proportion of people who would respond truthfully to the sensitive question. This is done with the help of RRT models. In the survey, the students were asked to respond to two binary randomization models. Model 1 is binary Greenberg et al. (1969) model that is used to estimate A , the proportion of people who would respond truthfully to a sensitive question. This is followed by Model 2 which is Binary Mixture Lovig et al. (2021) model and is used to estimate the prevalence of the sensitive trait which is having a 'troubled childhood'. In the survey, the randomisation device used for the binary model 1 consisted of 100 cards on which the sensitive question 'I trust the model' is written on 70 cards and the remaining 30 cards read the unrelated innocuous question 'My birthday falls in the month having 31 days'. Out of 484 responses, 364 'yes' responses were reported whereas 120 'no' responses were reported. The parameters chosen for Model 1 were $p_0 = 0.7, \pi_{y0} = 7/12$. Therefore, the estimated value of A is 0.8244 using (1).

The second randomisation device employed in the survey was based on Model 2 proposed by Lovig et al. (2021) using a deck of 100 cards. In the binary mixture RRT model of Lovig et al. (2021), the respondent randomly selects whether to answer regarding being in the direct question group, indirect question group, or unrelated question group. In the deck of 100 cards, 60 cards read 'I had a troubled childhood', 25 cards read 'I did not have a troubled childhood' and 15 cards read 'I was born in the month of July, August or September'. Thus, the values of parameters were chosen for Model 2 as $p = 0.6, q = 0.25, \pi_y = 3/12$. Out of 484 responses, 166 'yes' responses were received, giving us an estimate $\hat{\pi}_x$ of the prevalence of the sensitive characteristic as 0.1923 using (3).

It is clearly evident that the estimate of π is highest at 0.2665 through the anonymous response method, whereas, the estimate obtained using the RRT method is 0.1923. Furthermore, in a face-to-face survey, it turned out to be 0.1612 which is the least among these as expected.

It is worth noting that when the respondents were asked about the prevalence of sensitive characteristic through face-to-face method, then 16.12% of respondents reported having experienced troubled childhood and this percentage rose significantly by slightly over 10% when the respondents responded through complete anonymous method. This significant difference is due to the prevalence of social bias wherein the people wish to respond in a manner which puts them in a socially desirable setup. Hence, using the RRT model is better.

4. DISCUSSION AND CONCLUSION

The survey was aimed to validate the Binary Mixture Lovig et al. (2021) RRT model in a real scenario where the respondents were not likely to respond truthfully when confronted with a sensitive question. The respondents were requested to respond through three response techniques: Confidential, Face-to-Face and RRT. It was found that the RRT model in question aids in obtaining better approximation as compared to face-to-face and anonymous techniques. Additionally, it was observed that more than

82% of respondents had trust in the model and believed that their true response would remain private and confidential. It was also noted that the Face-to-Face technique yields the lowest estimate of the prevalence of troubled childhood, while the confidential method produces the highest estimate. As expected, the estimate obtained by RRT method lies between the estimates obtained from the Face-to-Face and Confidential methods. This survey attempt led to a better understanding of the need for careful implementation when conducting a survey using a complex model such as that of Lovig et al. (2021).

5. ACKNOWLEDGEMENT

The authors acknowledge Principal, Hindu College. This research project was carried out under the Dr. B. M. Bhatia Memorial Research Fellowship.

6. REFERENCES

- Chhabra, A., Dass, B. K., Gupta, S., 2016, *Estimating Prevalence of sexual abuse by an acquaintance with an optional unrelated question RRT model*. North Carolina Journal of Mathematics and Statistics **2**, 1–9.
- Downey, C., Crummy, A., 2022, *The Impact of Childhood Trauma on Children's Wellbeing and adult behavior*. European Journal of Trauma & Dissociation **6**(1), Article 100237.
- Gill, T. S., Tuck, A., Gupta, S., Crowe, M., Figueroa, J., 2013, *A field test of optional unrelated question randomized response models: Estimates of risky sexual behaviors*. In Topics from the 8th Annual UNCG Regional Mathematics and Statistics Conference, 135–146, Springer.
- 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**(326), 520–539.
- Lorenc, T., Lester, S., Sutcliffe, K., Stansfield, C., Thomas, J., 2020, *Interventions to support people exposed to adverse childhood experiences: systematic review of systematic reviews*. BMC Public Health **20**, 657. <https://doi.org/10.1186/s12889-020-08789-0>
- Lovig, M., Khalil, S., Rahman, S., Sapra, P., Gupta, S., 2021, *A mixture binary RRT model with a unified measure of privacy and efficiency*. Communications in Statistics: Simulation and Computation. <https://doi.org/10.1080/03610918.2021.1914092>
- Ostapczuk, M., Musch, J., Moshagen, M., 2009, *A randomized-response investigation of the education effect in attitudes towards foreigners*. European Journal of Social Psychology **39**(6), 920-931.
- Sihm, J. S., Chhabra, A., and Gupta, S. N., 2016, *An optional unrelated question RRT model*, Involve: A Journal of Mathematics, **9**(2), 195-209.
- Solomon, J., Jacobson, S. K., Wald, K. D. and Gavin, M. 2007, *Estimating illegal resource use at a Ugandan park with the randomized response technique*, Human Dimensions of Wildlife, **12**, 75–88.
- Tomlinson, T. A., Mears, D. P., Stults, B. J., Meldrum, R. C., Turanovic, J. J., Young, J. T. N., 2022, *The Legacy of Troubled Childhoods: Adverse Childhood Experiences, Sleep, and Delinquency*. Crime & Delinquency, 1-28. <https://doi.org/10.1177/00111287221083961>
- Warner, S.L., 1965, *Randomized response: A Survey Technique for Eliminating Evasive Answer Bias*. Journal of the American Statistical Association **60** (309), 63-69.