

Application of Randomized Response Techniques in Estimation of prevalence of Child Sexual Abuse

R.Srivastava¹, A.K. Nigam¹ and Neelam Singh²

¹*Institute of Applied Statistics and Development Studies, 1/172, Virat Khand, Gomtinagar, Lucknow-226010, U.P., India*

²*Vatsalaya, C-377, Church Road, Indira Nagar, Lucknow-226016, U.P., India*

Received February 18, 2015; Accepted April 02, 2015

Abstract

In the conduct of socio economic and the related surveys, the researchers are often faced with the problem of extracting accurate information on sensitive issues from the respondents. For sensitive questions the respondents are often reluctant to divulge true information. This leads to erroneous conclusions. In such a situation use of Randomized Response Technique (RRT) is recommended. RRT is a specially designed method to improve the accuracy of responses to sensitive questions. Through this technique interviewer is able to estimate the proportion of respondents associated with the 'sensitive question' without knowing the actual status of individual respondent. RRT has been used extensively in some countries to gather information on sensitive issues like the proportion of tax evaders in the country. The use of RRT, however, is rather scanty in India. This can fruitfully be employed to various fields like health, economics and issues related to social stigmas like sexual abuse etc. Child abuse is not only violation of their rights but it can also make them psychologically depressed and even place them in long-term psychological trauma. It is really difficult to get accurate responses due to stigma attached to it. We discuss the use of RRT to assess the extent of sexual abuse amongst children from data of a project carried out by Vatsalya. There were three categories of children who were either victims of 'severe sexual abuse', or 'mild sexual abuse' or 'no sexual abuse'. Abu'l-ela et. al. (1967) version of Warner (1965) for multi-proportions RR was used to estimate the proportions of sexually abused children.

Keywords: Randomized response technique; child sexual abuse.

1. Introduction

In the conduct of surveys related to socio economic and the related fields and health and nutritional aspects of children, researchers are often faced with the problem of extracting accurate information on sensitive issues like drug abuse, sexual behavior and tax evasion etc. Respondents are generally reluctant to respond truthfully to 'sensitive questions'. This leads to erroneous conclusions. In such a situation, use of Randomized Response Technique (RRT) is recommended which guarantees the privacy of the respondent. It allows the interviewer to gather the information more accurately. RRT allows estimation of the proportion of the population possessing a sensitive characteristic. The technique utilizes a randomized device

on a probability basis, which ensures confidentiality of the respondent. As the status of the respondent (whether he/she belongs to a sensitive group) is not revealed to the interviewer, the technique is likely to provide a much more reliable estimate, than the one obtained through direct questioning. The response in RRT is dichotomous (yes or no). The validity of RRT has been established in various studies by comparing the known estimates obtained by other methods and estimates obtained using RRT. The technique has been used extensively in some countries to gather information on sensitive issues like the proportion of tax evaders in the country. In India, however, its use appears to be rather scarce.

Child abuse is another area in which no reliable estimates are available in our country. Ministry of Women and Child Development, Govt. of India, in its report entitled 'Study on Child Abuse India 2007' has data on child sexual abuse obtained through direct questioning at National level in 5-18 year age group children. Child abuse is a violation of children's rights and can also place them in long-term psychological trauma. Other effects can include post-traumatic stress disorder, anxiety, (http://en.wikipedia.org/wiki/Child_sexual_abuse - cite_note-levitan-8) complex post-traumatic stress disorder, propensity to further victimization in adulthood, and physical injury to the child. Sexual abuse by a family member is a form of incest and can result in more serious and long-term psychological trauma, especially in the case of parental incest. It is really difficult to get accurate direct responses due to stigma attached to it. The present article demonstrates that RRT can be successfully used in the problem of child sexual abuse.

2. Preliminaries

Randomized response is a research method used in structured survey interview. Warner (1965) introduced this technique and it was later modified by Greenberg et al (1969). This technique allows respondents to respond to sensitive issues (such as criminal behavior or sexuality) while maintaining confidentiality. Chance decides, unknown to the interviewer, whether the question is to be answered truthfully, or "yes", regardless of the truth. In Warner's (1965) version, the sensitive question is worded in two dichotomous alternatives, and chance decides, unknown to the interviewer, which one is to be answered honestly. The interviewer receives a "yes" or "no" response without knowing to which of the two questions it belongs. Let p be the probability to answer the sensitive question and π the true proportion of those interviewed bearing the embarrassing property, then the proportion of "yes"-answers λ is composed as follows:

$$\lambda = p\pi + (1-p)(1-\pi)$$

Yielding

$$\pi = \frac{\lambda + p - 1}{2p - 1}$$

If the number of YES responses in a sample of size n is X , we estimate π with

$$\hat{\pi} = \frac{\frac{X}{n} + p - 1}{2p - 1}$$

The variance of the estimate of π is given by

$$V(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{p(1-p)}{n(2p-1)^2}$$

The ideas described above are illustrated through an example.

Example 2.1: This example relates to estimation of the proportion of respondents consuming cocaine. The respondents were given to respond to following alternative questions.

- Alternative 1: "I have consumed cocaine."
- Alternative 2: "I have never consumed cocaine."

The interviewees were asked to secretly throw a die and answer the first question only if they throw a 6, otherwise the second question ($p=1/6$). The "yes"-answers are now composed of consumers who have thrown a 6 and non-consumers who have thrown a different number. Let the result be 75 "yes"-answers out of 100 interviewed ($\lambda=3/4$). Inserted into the formula we get

$$\hat{\pi} = \frac{\frac{3}{4} + \frac{1}{6} - 1}{\frac{1}{3} - 1} = 1/8$$

This leads to the inference that if the respondents have answered truthfully then 12.5 % had consumed cocaine.

There are several other methods related to RRT that are more efficient. In this communication we however limit ourselves to Warner (1965) and its variants.

3. Some Earlier Studies

This section is devoted to some interesting case studies where RRT has been successfully employed. Savasan (2003) used RRT to model underground economy in Turkey. Abul-Ela et al 's (1967) RRT method and vector RR version (Bourke and Delenous, 1976) were used in this study. Unlike Warner (1965) version Abul-Ela et al (1967) use dichotomous response ('yes' or 'no') for polychotomous population. In this study under-reporting of income was divided into four categories viz.; 'less than 5%', 'between 6-25%', 'between 26%-50%' and 'more than 50%'. The respondents belonged to the fields of agriculture, industry-manufacturing and service. The advantage of vector RRT is that it is more efficient and requires smaller sample. For details see Savasan (2003).

A similar study was conducted by Houston and Tran (2000) on tax evasion in Australia using RRT and Direct Question (DQ) approach. Of the respondents completing the RRT, 5.5% admitted tax evasion by under-reporting income, and 6.5% admitted tax evasion by over-claiming deductions. The corresponding proportions obtained from the DQ survey instrument were 1.7% and 4.2% respectively.

Above case studies clearly illustrate the effectiveness of RRT if properly used as per the level of enquiry. RRT has been used extensively in other parts of the world, however, its use in India appears to be almost nil.

In India, the Institute of Applied Statistics and Development Studies (IASDS) in 1996 used RRT to assess the high risk sex behavior by highway truck operators in the state of Uttar Pradesh. The application of RRT was part of a project: *Targeted intervention among truck operators in Lucknow district to increase the levels of STD/HIV/AIDS information and awareness and distribution of condoms*. This World Bank funded project was sponsored by State AIDS Control Cell, UP Government in 1996. The study involved 606 (397 national, 209 domestic) truck operators. One of the objectives of the study was to estimate the proportion of truckers who belong to high risk category in terms of contracting HIV

infection. The study provided estimates of truckers belonging to high risk category, who had not been using condoms or were using it occasionally in their extra marital sexual activity. There seemed to be more enthusiasm among truckers in answering the questions relating to sexual behavior in spite of this being sensitive question related to promiscuous sex. This was because of use of RRT that uses indirect approach to seek the sensitive information. To answer the question the truckers were first asked to draw one of the slips at random which were of two colors – 7 blue and 3 red. They were chosen to answer ‘yes’ or ‘no’ to one of the two options as below depending upon the color of the slip drawn.

- I did not have sex with other than wife/ I had sex with other than wife but invariably used condoms.
- I had sex outside but did not always use condom.

The first option is indicative of safe sex whereas the second reflects risky sexual behavior. This procedure looked attractive to truckers because of confidentiality of their response. This, however, needed considerable explanation to how to respond in a randomized manner as required under RRT. After being convinced of confidentiality the truckers responded in an enthusiastic manner.

The RR response estimate of the second group of truckers after adjusting for few direct questions was 53 percent. The percentages of truckers belonging to high risk behavior for the National and UP/local categories were respectively 58 percent and 45 percent. Thus, large proportion of truckers was at the risk of contracting HIV infection due to unsafe sex.

4. The Present Study

The present write up is part of a project: *State level study on status of assessment of child sexual abuse in Uttar Pradesh*. The study financed by Plan India was carried out by Vatsalya in 2011. IASDS partnered in this study for aspects like formulation, analysis and reporting. The survey was carried out in Mau and Maharjganj (Eastern U.P), Lucknow and Ambedkarnagar (Central U.P.), Baghpat and Etawah (Western U.P.) and Jhansi and Jalaun (Bundelkhand) districts.

Among *various* child abuses, sexual abuse is very common. It is really difficult to get accurate responses due to stigma attached to it. In order to circumvent the problem use of RRT is desirable. Use of RRT, therefore, formed an important component of the study. Focus of the study was to assess the extent of sexual abuse in children in the age groups 5-10 years and 11-17 years. For children in the age group 5-10 years, information was gathered by direct questioning method. The children belonging to 11-17 years age group were classified into four categories viz.; (i) Children in family environment (ii) School going children (iii) Working Children and (iv) Children in institutional care. The coverage of these was 162, 167, 512 and 630 respectively. The category-wise breakup according to males was 72, 93, 223, 292 and that of females was 90, 74, 289 and 338 respectively.

Following indicators of sexual abuse were used.

Sexual Abuse (Severe Form)	Sexual Abuse (Mild Form)
<ul style="list-style-type: none"> ❖ Sexual assault. ❖ Make the child fondle private parts. ❖ Make the child exhibit private parts. ❖ Exhibit private parts to the child. ❖ Photograph the child in the nude. 	<ul style="list-style-type: none"> ❖ Forcible kissing, ❖ Forcible kissing during travel situations. ❖ Kissing/molestation attempt during marriage. ❖ Forced to view private body parts of others, ❖ Forced to view pornographic video/photographs

Information regarding sexual abuse in the children in the age group 11- 17 years belonging to family environment and working children was gathered using DQ method while for school going children and children in institutional care was obtained by using RRT. An overall sample of 512 (223 males, 289 females) for school going children and 630 (292 males, 338 females) for children in institutional care was used in the study. Further, the study was carried out in five zones of the state. RRT was applied in the following manner.

5. Methodology

Children (11-17years) were divided into two groups. Two sets of different urns were prepared with tokens of different colors as follows

	Urn1	Urn2
	4 (red)	3 (red)
	3 (Green)	4 (Green)
	3 (yellow)	3 (yellow)
Total	10	10

Group 1 was asked to pick a token from Urn1 and the following action was taken

Color of the token	Action
Red	Ask the child to respond YES/NO whether there was anytime in the past Sexual assault, or was made to fondle or exhibit private parts, or he/she exhibited his/her private parts, was photographed in the nude. Put the answer sheet in box 1
Green	Ask the child to respond YES/NO whether he/she was kissed forcibly, or some body tried to kiss/molest him/her during travel, or some body tried to kiss/molest him/her during marriage, or he/she was forced to view private body parts of others, or he/she was forced to view pornographic video/photographs Put the answer sheet in box 1
Yellow	Ask the child to respond YES/NO whether he/she was never sexually abused. Put the answer sheet in box 1

Similar procedure was carried out for Group 2 and answer sheets were collected in box 2.

The problem now is to estimate π_1, π_2, π_3 the true proportions of each statement,

$0 < \pi_j < 1$ and $\sum_{j=1}^3 \pi_j = 1$. Abul-Ela et al (1967) RRT version was utilized for this. .SRS

without replacement of sizes n_1 and n_2 were drawn independently from the population. For $i=1,2$ and $j=1,2,3$ P_{ij} denotes the proportion of cards belonging to j^{th} statement from i^{th}

sample. Then $\sum_{j=1}^3 P_{ij} = 1$ and the probability of 'yes' for any interviewee in i^{th} sample is

$$\lambda_i = \sum_{j=1}^3 P_{ij} \pi_j, \quad i = 1, 2$$

In our case

$$\lambda_1 = P_{11}\pi_1 + P_{12}\pi_2 + P_{13}\pi_3$$

$$\lambda_2 = P_{21}\pi_1 + P_{22}\pi_2 + P_{23}\pi_3$$

Remembering that $\sum_{j=1}^3 \pi_j = 1$ we have in matrix notation

$$\mathbf{P}\boldsymbol{\pi} = \boldsymbol{\xi}$$

Where

$$\mathbf{P} = \begin{pmatrix} P_{11} - P_{13} & P_{12} - P_{13} \\ P_{21} - P_{23} & P_{22} - P_{23} \end{pmatrix}$$

$$\boldsymbol{\pi} = \begin{pmatrix} \pi_1 \\ \pi_2 \end{pmatrix} \text{ and } \boldsymbol{\xi} = \begin{pmatrix} \lambda_1 - P_{13} \\ \lambda_2 - P_{23} \end{pmatrix}$$

If in the i^{th} sample n_{i1} persons report 'yes' then an unbiased estimate of λ_i is $\hat{\lambda}_i = n_{i1} / n_i$.

There fore writing $\mathbf{c} = \begin{pmatrix} \hat{\lambda}_1 - P_{13} \\ \hat{\lambda}_2 - P_{23} \end{pmatrix}$ we have

$\hat{\boldsymbol{\pi}} = \mathbf{P}^{-1}\mathbf{c}$ provided inverse exists: we may otherwise use Moore-Penrose inverse. Since π_1 and π_2 have been estimated, π_3 can easily be estimated.

Dispersion matrix of \mathbf{c} is given by $\text{disp}(\mathbf{c}) = \text{diag}(V_{11}, V_{22}, V_{33})$ and dispersion matrix of $\boldsymbol{\pi}$ is given by $\text{disp}(\boldsymbol{\pi}) = \mathbf{P}^{-1} \text{diag}(V_{11}, V_{22}, V_{33})$ and $V_{ii} = \lambda_i(1 - \lambda_i) / n_i$

In our case

$$P_{11} = P_{22} = 4/10$$

$$P_{12} = P_{13} = P_{21} = P_{23} = 3/10$$

The estimating equations are

$$\hat{\pi}_1 = 10\hat{\lambda}_1 - 3,$$

$$\hat{\pi}_2 = 10\hat{\lambda}_2 - 3;$$

$$\hat{\pi}_3 = 1 - (\hat{\lambda}_1 + \hat{\lambda}_2)$$

where

π_1 = Proportion of severe sexual abuse

π_2 = Proportion of mild sexual abuse

π_3 = Proportion of no sexual abuse

λ_1 = Proportion of yes response in group 1

λ_2 = Proportion of yes response in group 2

Limitations

It can be seen from estimating equations that following should hold under the above set up

$$0.3 < \lambda_1, \lambda_2 < 0.4$$

and

$$0.6 < \lambda_1 + \lambda_2 < 0.7$$

Where-ever these two conditions do not hold the estimation through Randomized Response Technique by subgroups is not possible.

This in simple terms means that proportions of 'yes' responses in individual groups should be between 30 percent 40 percent and when two groups are taken together the total of 'yes' responses should be between 60 percent to 70 percent. This exercise yielded the following results

6. Sexual Abuse across the zones

Table 2 provides a comparative scenario of incidence of overall child sexual abuse across various categories of children. It may be observed randomized response technique has been able to capture higher incidence of abuse among children of *school going* and *institutional care* categories with the latter category having higher percentages. A possible reason of this could be relatively higher exposure to outsiders for 'children in institutional care' category children due to residential nature of the institution. However, one should also take note of the fact that children of Family environment spend more time at home and are more protected.

Zone	Category			
	Family environment	Working	School going*	In institutional care*
Eastern U.P.	64.1	61.3	76.4	94.6
Central U.P.	45.5	53.8	75.0	90.4
Western U.P.	71.1	65.1	76.8	86.1
Bundelkhand	36.6	41.5	91.1	84.2
U.P. Overall	53.7	55.1	80.2	88.7

* *Estimated using Randomized Response Technique*

Table 3 relates to severity of sexual abuse. Here the reporting gives higher percentages (over 20%) in all categories of children who were put to direct questioning. However, the children who were administered Randomized Response technique had lower percentages of

severe abuse. It could be because of this technique capturing the two types of abuse more correctly and because of confounding (mix-up) between severe and mild abuses when the questions are put directly.

Type of sexual abuse (%)	5-10 yrs	11-17 yrs			
		Family environment	Working	School going*	In institutional care*
Sexual Abuse	41.1	53.7	55.1	80.2	88.7
Severe Sexual Abuse	22.7	24.1	27.6	16.6	18.4

* *Estimated using Randomized Response Technique*

Ministry of Women and Child Development, Govt. of India, in its report entitled 'Study on Child Abuse India 2007' has provided information (Table 4) on child sexual abuse relating to above categories at National level. This study has considered children in the age group 5-18 years, while our study has two classifications viz; 5-10 years and 11-17 years. Thus a comparison between the two studies will be a pointer only i.e., a rough comparison. Comparison with figures in Table 2 shows that the estimates obtained using RRT (i.e., School going Children and Children in institutional care in our study are fairly high as compared to study carried out by the Ministry of Women and Child Development, clearly establishing the advantage of using RRT.

Family environment	Working	School going	Institutional Care
53.18	61.61	49.92	47.08

Conclusion: This study has clearly established the advantage of use of RRT in sensitive issues like sexual abuse where getting accurate responses is very difficult. RRT can be very successfully used in various other related fields of health and nutrition where sensitive issues are involved.

Acknowledgement: The authors thankfully acknowledge the financial support received from PLAN India to carry out the study entitled 'State Level Study on Status of Assessment of Child Sexual Abuse in Uttar Pradesh (2012).

References

- Abul Ela, Abdel Latif, A., Greenberg, B.G., and Horvitz, D.G. (1967). A multi-proportions RR model. *Journal of American Statistical Association*, **62**, 990-1008.
- Bourke, R.D. and Dalenius, T. (1976). Some new ideas in the realm of Randomized Inquiries. *International Statistics Review*, **44**, 219-221
- Greenberg, B., Abdul Ela, A., Simmons, W., and Horvitz, D. (1969). The unrelated question randomized response model: theoretical framework. *Journal of American Statistical Association*, **64(326)**, 520-39.
- Houston, J. and Tran, A. (2000). A Survey of Tax Evasion Using the Randomized Response Technique. *Paper presented in Tax Research Network Conference in Birmingham, the United Kingdom.*

- Savasan, F. (2003). Modeling the Underground Economy in Turkey: Randomized Response and MIMIC Models. *Bu makale the Journal of Economics*, **XXIV**(1), 47-76.
- State Level Study on Status of Assessment of Child Sexual Abuse in Uttar Pradesh (2012). *IASDS and Vtsalaya study funded by PLAN India*
- Study on Child Abuse India 2007 (2007). *Ministry of Women and Child Development, Govt. of India*
- Targeted Intervention among Truck Operators in Lucknow district to increase the levels of STD/HIV/AIDS information and awareness and distribution of Condoms (1996): *Institute of Applied Statistics and Development Studies, Lucknow.*
- Warner, S.L. (1965). Randomized response – a survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, **60**, 63-69.