NCAER | SINCE 1956
QUALITY. RELEVANCE. IMPACT

NCAER NATIONAL DATA INNOVATION CENTER

MEASUREMENT BRIEF 2024

NDIC FELLOWS PROGRAMME

Social Desirability and Interviewer Bias on Responses to Sensitive Questions in Large-Scale Surveys

SAURABH SINGH

Overview and measurement challenges

Surveys have been a vital source of information on the demographic and health characteristics of a population, primarily because they are not time-consuming and are cost-effective. However, there is always a high risk of measurement errors or mismatches between actual circumstances and survey responses, particularly when sensitive questions are involved. Such questions often lead to higher non-responses or socially acceptable responses that lack validity and reliability. Sensitive/emotional and attitudinal questions are more likely to result in varied responses (Schaeffer et al., 2010). In addition, interviewer characteristics such as gender, race, age, and previous interviewing experience can influence the response to a sensitive survey question (Singh et al., 2024). The effects of the interviewer on sensitive questions were found much larger in the fourth wave of the National Family Heath Survey (NFHS) than in the third wave (Singh et al., 2022).

In India, surveys have tried to capture accurate estimates on topics such as substance use, abortion, reproductive and sexual behaviour, and violence, but these are sensitive topics primarily due to widespread taboos and a male-dominated society. This study aims to assess the effect of interviewer characteristics on the response to sensitive questions about substance abuse at four levels—individual, interviewer, primary sampling unit (PSU), and district—using data from the fifth wave of the National Family Health Survey (NFHS-5).

Method and Data

The study uses the data from NFHS-5, a nationally representative cross-sectional study conducted under the Ministry of Health and Family Welfare (MoHFW), Government of India. The Fieldworker Characteristics file (FW file) and Individual file (IR file) were used for the analysis. Fieldworker's data file provides information about interviewer's characteristics that they filled via a self-administered questionnaire before the fieldwork began. Duplicate fieldworkers were dropped (2,010 out of 7,127) and only data on female fieldworkers (2,950) was used. The final sample for the study comprises 42,247 women from the NFHS state module after dropping unmarried women.

Dependent Variables
The dependent variables are an individual's alcohol and tobacco consumption. Respondents were asked if they asked if they drink alcohol and if they use tobacco in any form.

Explanatory Variable
The explanatory variable is interruption of the interview. The interviewer reported the presence of husband, a female adult or a male adult during the interview.

Independent Variables
There are two sets of independent variables—one for the interviewer and the other for respondents.

Interviewer characteristics include place of residence, age, children ever born, education, religion, caste, experience of NFHS/ other surveys, and mother tongue of the interviewer were used. Additionally, the number of interviews taken by the interviewer) during the entire survey period was clubbed into three quantiles (low with a mean of 16.6 interviews, medium with a mean of 28.4 interviews and high with a mean of 41.3 interviews).

Respondent characteristics include age, level of education, occupation, exposure to mass media (combining reading newspapers, listening to radio, and watching television) and wealth quintile. Other variables were caste, religion, place of residence, and region (East, West, North, South, North-East, and Central). The respondent's time of interview was grouped into three periods of the survey: Phase 1, Phase 2 Pre-covid and Phase 3 post-Covid.

Statistical Analysis

Univariate Analysis was used to document the interruption of the interview. Bivariate Analysis was used to show the reporting of the dependent variables in different scenarios of interruption. Additionally, a Chi2 test was performed to check the significance of this reporting. Bayesian cross-classified random intercept multi-level logit estimation method was used to examine the interviewers' effect on selected outcome

variables by adjusting all the independent variables. Further, a sensitivity analysis was done with other more sensitive and less sensitive questions asked in the survey. The detailed estimation method can be found in Singh et al. (2024).

The coefficient plot obtained from the logistic analysis was used to assess the effect of interviewer characteristics along with the interruption of the interview. Further, propensity score matching was used to assess the effect of treatment (here, interruption) on the reporting of the dependent variables.

Results

Figure 1 suggests that around 17 % of the interviews were interrupted by the third person, of which around 65% of the interviews were interrupted by all (spouse, male and female adults). However, the survey did not provide details of the interruption such as timing and duration of the interruption, or type of interruption (formal or informal). Figure 1 also shows that around 21% of interviews were interrupted by the husband only, followed by a female adult (4.6%), and 3.6% of the interviews were interrupted by the husband and another male adult. So, it is evident that around 30% of the interrupted interviews were interrupted by the husband.

Table 1 shows the unadjusted odds of how interruption affects the reporting of alcohol and tobacco consumption. Alcohol consumption is likely to be reported less when the interruption occurs more than once in comparison to no presence/interruption during the interview. The reporting of smoking behaviour is

Figure 1: Interrupted Interview by Third Person

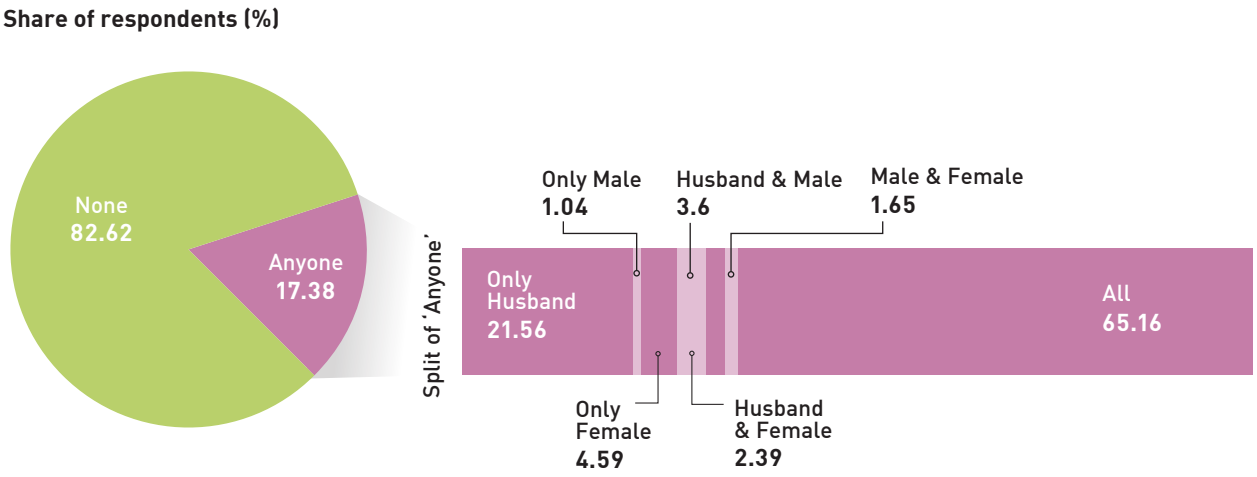


Table 1: Presence of the third person affects the Reporting of Sensitive Questions

Husband Presence	Alcohol	Smoke	Husband Alcohol
Once	1.1(0.9-1.36)	1.07(0.94-1.21)	1.01(0.94-1.1)
More Than Once	0.75**(0.6-0.95)	0.85**(0.74-0.96)	0.65***(0.6-0.71)
Any Presence	0.92(0.78-1.07)	0.95(0.86-1.04)	0.81***(0.76-0.86)
Male Presence			
Once	0.88(0.67-1.18)	0.99(0.84-1.16)	0.89**(0.8-0.99)
More Than Once	0.63***(0.49-0.81)	0.89*(0.78-1.02)	0.59***(0.54-0.65)
Any Presence	0.72***(0.59-0.88)	0.93(0.83-1.03)	0.7***(0.65-0.75
Female Presence			
Once	0.83(0.63-1.09)	0.99(0.84-1.15)	0.88**(0.8-0.97)
More Than Once	0.61***(0.47-0.79)	0.95(0.83-1.08)	0.63***(0.58-0.69)
Any Presence	0.7***(0.57-0.85)	0.96(0.87-1.07)	0.73***(0.68-0.78
Anyone Presence			
Anyone	0.89(0.76-1.04)	1(0.92-1.1)	0.83***(0.79-0.88)

around 15%, and 11% lower when the husband and any male interrupted the interview more than once. But when the question is about the husband's alcohol consumption, it is statistically significantly lower, irrespective of whether the interview was interrupted once or more than once.

Figure 2 (on Page 5) shows the coefficient plot obtained from logistic regression analysis. The results indicate that the reporting of smoking and alcohol consumption among respondents and the husband's alcohol consumption was lower when interview was interrupted by someone and the interviewer shared the same language. However, the reporting of alcohol consumption was significantly higher when interviewer's residence was a town or rural and the interviewer was older than the respondent. The reporting of husband's alcohol consumption was significantly higher when the interviewer had the same place of residence, and when the interviewer was older and resided in the same state.

Figure 3 (on Page 5) shows the variance at different levels, i.e., District, PSU, Interviewer, and Individual, explaining the contribution of each level on the dependent variables For sensitive behaviours such as respondent alcohol use, smoking, and husband's alcohol use, the interviewer level accounts for the majority of the variance (11% to 22%), indicating that these behaviours are primarily influenced by individual characteristics and then by interviewer characteristics rather than clustering effects. In contrast, more sensitive variables related to domestic violence (e.g., father beating mother, partner physically hurt, bruises caused by husband) show substantial variation at the Interviewer level than in the case of a less sensitive question by approximately 21%.

Policy Lessons

The fifth wave of the National Family Health Survey (NFHS-5) provides critical insights into how interviewer characteristics and interview interruptions impact the reporting of sensitive behaviours such as alcohol and tobacco use. Accurate self-reported data on these behaviours is essential for public health policy and intervention strategies, particularly in areas involving sensitive behaviours such as substance use. The findings of this study offer valuable insights into bias in data collected on sensitive topics.

Fear of judgment and reluctance to disclose it to a stranger can affect the likelihood of respondents reporting alcohol and substance use. To reduce the impact of interruptions, it is crucial to enhance training programmes for interviewers. Training should focus on techniques to manage and minimise interruptions, such as establishing clear interview boundaries and engaging with family members to ensure uninterrupted sessions. Conducting interviews in private or semi-private settings can also help reduce the likelihood of interruptions, particularly from household members.

These results show that even though it may not be possible to match interviewers and respondents on all important characteristics, the influence of interviewer-respondent dyad characteristics should be assessed for analysis on subjects with strong social role expectations.

Fieldworker deployment should be strategic to minimise biases in reporting sensitive behaviours. Matching interviewers and respondents by age and gender might be an effective way to reduce interviewer variance, as could supporting interviewers as they cope with their own emotions after hearing about other's adversity. Additionally, assigning older interviewers and those familiar with rural or town environments may enhance the accuracy of reporting regarding substance use.

The use of ACASI and C-ACASI (Color-coded ACASI) has showed potential for gathering private information from uneducated men and women (Bhatnagar et al., 2013) by providing women more comfort in answering sensitive questions and thus lowering underreporting.

Developing and implementing standardised protocols can guide how to handle interruptions and other potential disruptions. These protocols should include scripts for interviewers on addressing interruptions and maintaining the flow of the interview. Systematic documentation of interruptions, including their timing, duration, and context, is also essential to better understand their impact and refine future survey methodologies.

Figure 2: Factors Affecting Reporting of Sensitive Questions in NFHS-5, 2019-21

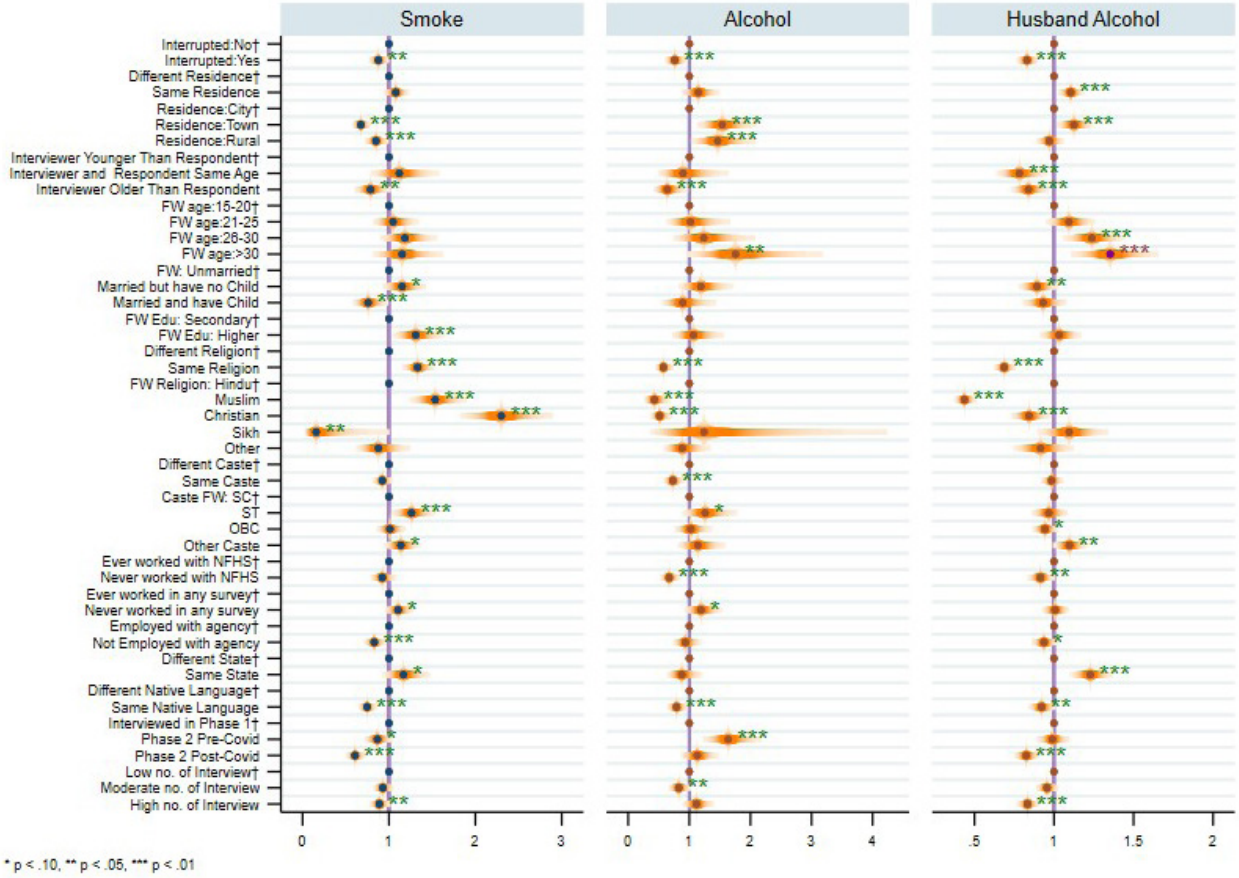
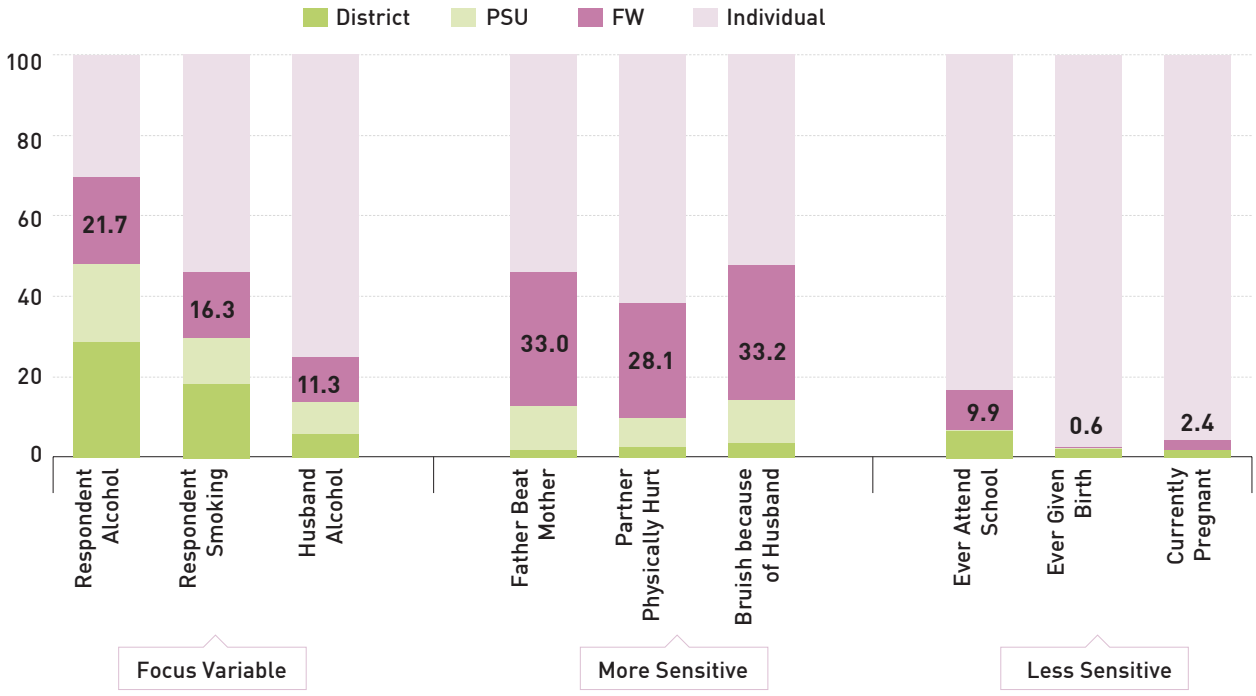


Figure 3: Percentage of estimated variance for selected outcomes explained by district, PSU and Interviewer level from cross-classified multilevel model



KEY FINDINGS

Impact of Interview Interruptions

The analysis indicates that approximately 17% of interviews experienced interruptions, with a significant portion (65%) involving all family members (spouse, male and female adults). Notably, the husband alone accounted for around 30% of interruptions, either solely or in conjunction with other adults. These interruptions had a measurable impact on the reporting of sensitive behaviours. For instance, the likelihood of reporting smoking behaviour decreased by 15% when the husband interrupted more than once, and by 11% when any male interrupted more than once.

Interviewer Characteristics

The analysis also indicates the influence of interviewer characteristics on the reporting of substance use. Reporting of smoking and alcohol use was lower when the interviewer shared the same native language as the respondent. Conversely, alcohol consumption reporting increased when the interviewer was older than the respondent and when the interviewer belongs to a town or rural area. Reporting of the husband's alcohol consumption was higher when the interviewer shared the same place of residence, was older, and resided in the same state. These results suggest that certain interviewer attributes can either facilitate or hinder the disclosure of sensitive information.

Older and more experienced interviewers are linked to higher reporting. Also, prior NFHS survey experience is associated with higher reporting. This study finds that those interviewers who have never worked with NFHS were statistically significantly less likely to report alcohol and substance use.

The interviewer effect was more pronounced for sensitive topics than in case of less sensitive questions used in the study. Interviewer effects were typically more pronounced in than District and PSU (primary sampling unit,) effects, which are due to random variations between districts or PSUs respectively.

References

Bhatnagar, T., Brown, J., Saravanamurthy, P. S., Kumar, R. M., & Detels, R. (2013). Color-coded audio computer-assisted self-interviews (C-ACASI) for poorly educated men and women in a semi-rural area of South India: "good, scary and thrilling". *AIDS and behavior,* 17(6), 2260–2268. https://doi.org/10.1007/s10461-013-0414-3

Bradley, S. E. (2016). When Quality Matters: Linking the Reliability of Demographic and Health Survey Data to Biases in International Mortality, Fertility, and Family Planning Estimates. Unpublished Doctoral dissertation, UC Berkeley.

Pullum, W. P., Juan C., Khan N., and Staveteig, S. (2018). The Effect of Interviewer Characteristics on Data Quality in DHS Surveys. DHS Methodological Reports No. 24. Rockville, MD: ICF.

Schaeffer, N. C., J. Dykema, and D. W. Maynard (2010), "Interviewers and Interviewing," in Handbook of Survey Research, Second Edition, eds. J. D. Wright and P. V. Marsden, Bingley, UK: Emerald Group Publishing Limited, 437–470.

Singh, A., Kumar, K., & Arnold, F. (2022). How Interviewers Affect Responses to Sensitive Questions on the Justification for Wife Beating, the Refusal to have Conjugal Sex, and Domestic Violence in India.

Studies in family planning, 53(2), 259–279. https://doi.org/10.1111/sifp.12193

Singh, S., Dwivedi, L. K., & Jana, S. (2023). Field-worker effects on quality of data collected on sensitive questions in National Family Health Surveys: An analysis based on intimate partner violence in India. *SSM - population health, 25,* 101557. https://doi.org/10.1016/j.ssmph.2023.101557

Suggested citation

Singh, Saurabh (2024). "Social Desirability and Interviewer Bias on Responses to Sensitive Questions in Large-Scale Surveys", Measurement Brief - NDIC Fellows Programme (2023-24). NCAER National Data Innovation Centre, New Delhi.

Author Bio

Saurabh Singh is a research scholar at the Department of Survey Research and Data Analytics, International Institute for Population Sciences (IIPS)-Mumbai. His Ph.D. research focuses on "Comparative Assessment of Data Collection Methods for Sensitive Questions in Indian Context." He has a bachelor's degree in Mathematics and Statistics (2014-17), a dual Master's degree in Health Statistics (2017-19) and Population Studies (2019-20), and M.Phil. in Population Studies (2020-21). He is proficient in statistical tools and software such as STATA, SPSS, Python, and Power BI and has hands-on experience handling extensive survey data like NFHS, LASI, and SAGE. His areas of interest are Biostatistics, Epidemiology, Demography, and Survey Research.

Funding Support and Disclaimer

This research is supported by the National Council of Applied Economic Research through the NCAER National Data Innovation Centre. The views presented in this paper are those of the authors and not those of NCAER or its Governing Body. Funding for the NCAER National Data Innovation Centre is provided by Bill & Melinda Gates Foundation.

About NCAER NDIC

The NCAER National Data Innovation Centre was set up in December 2017 to promote innovation and excellence in data collection and build research capacity to strengthen the data ecosystem in India. The NDIC is envisaged as a hub for providing expertise to policymakers, government statistical agencies and private data collection agencies. NDIC is involved in pursuing three primary activities:

1. To pilot innovative data collection methods and mainstream successful pilots into larger data collection efforts;
2. To impart formal and informal training to a new generation of data scientists; and
3. To serve as a resource for data stakeholders, including Government data agencies and ministries.

About NCAER

Established in 1956, NCAER is India's oldest and largest independent, non-profit, economic policy research institute. NCAER's work cuts across many sectors, including growth, macro, trade, infrastructure, logistics, labour, urban, agriculture and rural development, human development, poverty, and consumers. The focus of NCAER's work is on generating and analysing empirical evidence to support and inform policy choices. It is also one of a handful of think tanks globally that combine rigorous analysis and policy outreach with deep data collection capabilities, especially for household surveys. More on NCAER is available on www.ncaer.org.

Designed by: How India Lives (www.howindialives.com)

NATIONAL COUNCIL OF APPLIED ECONOMIC RESEARCH
NCAER India Centre, 11 Indraprastha Estate, New Delhi 110 002 (INDIA)
Tel: +91-11-2345 2657, 6120 2698, Email: info@ncaer.org
www.ncaer.org