

Interviewer Characteristics Related to Field Interviewer Falsification

Angel Griffin, James Pratt, Sean Shuangquan Liu, Lee Ellen Coffey, Stephanie A. Terrey, and Christine Clark

RTI International

May 2011

Send all correspondence to:

Angel Griffin

720 Fishing Rock Road

Castalia, NC 27816

Voice: (919) 316-3988

E-mail: agriffin@rti.org

Draft Version-Please do not quote without author permission.

ABSTRACT

Falsification of data is detrimental to the integrity of any research. When a field interviewer falsifies data costs increase when investigating possible falsification, reworking fraudulent cases, and hiring replacement staff. Given the negative impact of falsification on research integrity and budgets, survey data collection organizations and their sponsors have an interest in finding means of detecting and preventing falsification. Can we identify any interviewer characteristics that are directly associated with falsification? Do factors such as interviewer age, experience level, or other typically measured performance ratings such as response rate or cost per interview correlate with falsification? Are there differences between interviewers whose work is questionable but not falsified? Using the National Survey on Drug Use and Health (NSDUH) as an example, these questions will be addressed.

The National Survey on Drug Use and Health (NSDUH) is an annual face-to-face, household survey. The NSDUH is sponsored by the Substance Abuse and Mental Health Services Administration. First conducted in 1971, this study provides national, state and substate data on substance use and mental health in the civilian, noninstitutionalized population age 12 and older. Approximately 140,000 household screenings and 67,500 NSDUH interviews are completed annually and data are collected in all 50 states plus the District of Columbia by approximately 700 interviewers.

The NSDUH implements a complex process to verify the quality and accuracy of each interviewer's work. In some instances, an interviewer's work is field verified to determine whether or not he or she made proper contact. Using 2005 to 2010 non-identifiable archived data, this presentation will examine the correlation of detected falsification with factors such as interviewer experience on the NSDUH (considering both time and workload), age, and response rate history. Trends in the number of interviewers found falsifying and their characteristics will also be investigated as well as any differences between interviewers with questionable work requiring field verification, interviewers who falsified, and those who did not falsify.

Interviewer Characteristics Related to Field Interviewer Falsification

INTRODUCTION

Falsification is defined as “occurring whenever the interviewer knowingly deviates from current interviewing procedures to avoid interviewing, classifying, and/or listing units” (Schreiner et al., 1988). Falsification of data can lead to significant costs for any research project, both in terms of data quality and budget. Falsified data is the antithesis of quality data and impacts research integrity. Financially, when an interviewer falsifies data, costs increase when investigating possible falsification, reworking fraudulent cases, and hiring replacement staff. In field surveys, it is more challenging to prevent this intentional deviation since interviewers work independently in the field with limited supervision from project management.

Since falsification is not a new phenomenon, researchers have examined various approaches to detecting falsified data. Often researchers look at interview data to detect falsification (Bredl et al., 2008; Porras and English, 2004; Murphy et al., 2004). Some researchers use Benford’s Law, the distribution that leading digits of numbers should follow, to identify possibly falsified data (Judge and Schechter, 2009; Schrapler and Wagner, 2003; Swanson et al., 2003).

In addition to data-driven approaches to detecting falsification, Schreiner et. al (1988) investigated interviewer characteristics of falsification and found that interviewers with a shorter length of service were more likely to falsify. In our review of the literature, other factors such as cost per interview were not examined in regards to detecting falsification.

On the National Survey on Drug Use and Health (NSDUH), approximately 700 FIs are dispersed throughout the nation, working individually in their assigned areas. Like many other federally funded surveys, NSDUH has meticulous verification procedures in place to ensure data quality. Although falsification is rare on NSDUH, it still occurs and threatens data quality. Given that less experienced interviewers are more likely to falsify based on earlier research, are other interviewer characteristics like cost per interview associated with falsification as well?

This paper is a descriptive analysis focusing on differences between response and error rates as well as cost per interview among a group of interviewers with questionable work. Given that experience is associated

with falsification, this investigation will compare falsifiers to nonfalsifiers on the aforementioned factors among less experienced FIs and more experienced FIs.

BACKGROUND

The National Survey on Drug Use and Health (NSDUH) provides national, state and substate data on substance use and mental health in the noninstitutionalized civilian population of the United States age 12 and older. RTI International¹ (RTI) conducts data collection operations under contract with the Substance Abuse and Mental Health Services Administration (SAMHSA). A team of about 700 field interviewers (FIs) visit approximately 140,000 households and interview 67,500 persons each year, with the annual sample divided and worked on a quarterly basis. FIs complete a 5 minute screening with an adult resident of the household to obtain demographic information about the residents of the household age 12 and older. One, two, or no eligible members of a household could be selected for the interview which takes about an hour.

As part of a commitment to research integrity, RTI's standard practice concerning falsification has three pillars: deterrence, detection, and remediation. The first pillar of deterrence involves informing interviewers that falsification is not tolerated, that certain methods will be employed to detect falsification, and that, if found, falsification is grounds for termination. The second pillar of detection requires projects to determine the most appropriate method or methods for detecting falsification. These methods could include recontacting respondents to ensure data collection took place as well as analyzing available data to identify suspicious patterns. The final pillar of remediation involves reporting and correcting falsification. It is a current best practice for projects to attempt to repair falsification damage when possible and to document various details about the falsification.

NSDUH adheres to the procedures of deterrence, detection and remediation. FIs are informed at training about falsification and its ramifications. This information is also discussed in their field manual. All FIs acknowledge they are fully aware of the NSDUH's falsification policy by signing a formal agreement during training.

NSDUH uses telephone verification as a recontact method to verify the quality of FIs' work and to detect falsification. Respondents are asked to provide a phone number for verification at the completion of all

¹ RTI International is a trade name of Research Triangle Institute.

screenings for which no one in the household is eligible or selected for the interview. Respondents selected for the interview are asked to provide their phone number and address at the end of the interview. All respondent contact information is kept separate from the responses to interview questions and stored in the verification system. A percentage of each FI's work is selected for verification and called by trained telephone interviewers to verify that the FI followed project protocols. For interview cases without an available phone number but with a complete mailing address, a verification letter is sent to the respondent to complete and return by mail to RTI. The completed verification letters are then keyed and the results included in verification reports.

There are many potential barriers to actually completing verification calls with respondents. FIs may report that respondents refused to provide verification information or do not have a phone number, the number provided may not be a working number, or it may be the wrong number for the household. While a few incidences of these situations are expected, a high percentage of refusals, no phones, non-working and/or incorrect numbers can seriously impact the ability to monitor an interviewer's work and may raise concerns about the validity of the data collected. Other reports monitoring interview length, FI time reports, duplicate phone numbers, and discrepancies between screenings and interviews can also raise concerns about the validity of an FI's work.

When a significant percentage of an FI's work cannot be verified via telephone verification and/or significant problems with an FI's work are identified, a field verification is conducted. Field verifications involve training an experienced FI to return to selected addresses reportedly contacted by the original FI to verify the quality of work and data collected. Field verifications can become very costly and are used as a last resort when regular telephone verification procedures are unable to verify data quality. If falsification is discovered, the FI is terminated and the invalid cases are reworked.

Prior research into detecting interviewer falsification patterns on the NSDUH (Murphy, et al., 2004) analyzed available timing data, item nonresponse data, and interview data. Various data quality monitoring programs existed on the NSDUH; researchers instead focused on additional analysis of available data to enhance the quality systems and falsification detection measures already in place. Comparing state averages with the results of known falsifiers, researchers examined timing data, item nonresponse data, and select individual interview response data, looking for patterns of rare response combinations and deviation from normal

response patterns. Based on the research findings, several of these tests were implemented on the study beginning in 2004.

This prior research allowed for enhancements and increased vigilance on NSDUH in identifying those interviewers who may be submitting falsified work. Data quality team members monitor these FIs more closely, conducting increased verifications of the FIs' work and tracking other data quality measures. The data quality staff also partner with the supervisory staff to monitor various aspects of the interviewers' data collection activities.

METHODS

The NSDUH data are collected, stored and archived on a quarterly basis. Since 2005, NSDUH data collection procedures and strategies have remained consistent which permits comparisons of falsification data across survey years 2005 through 2010.

An FI selected for field verification must stop working and return all project equipment to RTI. Data retrieved for each FI involved in this investigation includes all case events entered, timesheets completed, and field verification information.

Between January of 2005 and December of 2010, 174 FIs were field verified of which 89 were found to have falsified data. A total of 2,094 different FIs, including those field verified, worked on NSDUH during this same time frame. This investigation looked at the following interviewer characteristics: workload, experience, response rates, cost per interview, and data quality error rates of field verified FIs.

Workload: Workload is a function of the number of cases worked where a case is an attempted screening or interview. For workload we counted the unique cases for which an FI entered at least one event during the quarter. The number for workload is essentially the number of cases the FI was assigned and subsequently began working.

Experience: Experience was measured by a count of weeks an FI worked on the project as determined by electronic timesheet submissions approved by a supervisor. The count consists of the total number of weeks an FI had an approved timesheet up to the point of being field verified. In the analysis, less experienced FIs are defined as having worked fewer than 52 weeks. More experienced FIs worked for 52 weeks or more. There were 112 less experienced FIs and 62 more experienced FIs in the sample.

Response Rates: The NSDUH maintains a record of screening response rate and interview response rate information to measure the performance of every FI on the project.

The screening response rate is the ratio of completed screenings to the number of potential screenings that could have been completed taking into account factors that would make a household ineligible for being screened such as being vacant.

The interview response rate is the ratio of completed interviews to potentially completed interview cases. Once a household is screened as a 1 interview or 2 interview household it is considered to be a potentially completed interview case.

Cost Per Interview: Each FI is required to submit a summary report every week they work that includes the total number of reimbursable miles traveled, any miscellaneous expenses, and hours worked. Supervisory staff closely monitor the summary information because the project uses the data for cost projection, but the summary data isn't always accurate or complete and is not used for interviewer payment. If an FI misses a week, that summary information cannot be added at a later date. For this reason there are some anomalies in our numbers. For example several FIs have an average miles traveled per quarter for an interview as zero due to the fact that an FI may have only worked for a short period of time and during that time they may have only submitted one summary of their work. If the FI didn't submit any miles traveled in that summary it is counted as 0. There were several FIs in this category and also in a similar situation for hours worked.

This summary report information was used to calculate the cost per interview for each FI per quarter. Total miscellaneous expenses, miles traveled, and hours worked were divided by the total number of completed interviews for the FI per quarter. This gave us averages for miscellaneous expenses per interview, miles per interview, and hours per interview for each FI for each quarter. Expenses per interview and hours per interview were then classified into below average, average, and above average categories using the mean value and one standard deviation of the field verified FIs.

DQ Error Rates: Every case chosen for field verification is assigned a final code, falling into one of three categories – Falsified, Questionable but not falsified (Inconclusive), or OK. If at least one case is found to be falsified, the FI who originally completed the case is considered to be a falsifier.

In this investigation we looked at the rate of screening and interview errors for each field verified FI and compared those with the average error rates of other FIs in the same state. Some of the factors used to

calculate the error rates are missing screening data, incorrectly completed paperwork, and incorrectly coded events. We divide the total number of interview and screening errors by the total number of interview and screening events, respectively, to calculate interview and screening error ratios for each FI.

Statistical Analyses:

Difference of means tests were used to examine differences between falsifiers and nonfalsifiers on our measures of interviewer characteristics including workload, response rates, miscellaneous expenses per interview, hours per interview, reimbursable miles per interview, and error rates. We carried these tests out separately among less experienced FIs and more experienced FIs. Dividing this analysis by experience allowed us to control for the fact that some of the above factors are associated with experience. Due to the proprietary nature of cost information, we do not report on the exact values of miscellaneous expenses per interview and hours per interview and instead use our categories of “above average”, “average” and “below average”. However, the tests of significance for these variables are based on the underlying quantitative data.

RESULTS/DISCUSSION

Difference of Means Tests Between Falsifiers and Nonfalsifiers Among Less Experienced Interviewers:

This investigation compared the workload, response rates, cost per interview, and error rates of falsifiers to nonfalsifiers among less experienced FIs. There were 112 FIs who worked on the NSDUH for less than 52 weeks. Of those 112 FIs, 65 (58 percent) were found to have falsified while the remaining 47 had questionable work but no clear evidence of falsification. Sixty-two FIs worked on the NSDUH for 52 weeks or more prior to being field verified. Twenty-four FIs (38.7 percent) were found to have falsified while the remaining 38 had questionable work but no clear evidence of falsification. This relationship between experience and falsification may just be an artifact of having a good detection system in place on the NSDUH. Falsifiers are caught relatively early in their careers and terminated so they do not get to become experienced interviewers.

The results for the difference of means test among less experienced FIs are displayed in Table 1. There was a significant ($P\text{-value} < 0.10$) difference in screening error rate between falsifiers and nonfalsifiers. The screening error rate ($P\text{-value} = 0.0732$) was higher among falsifiers indicating more data quality errors among that group. The difference between miscellaneous expenses per interview was very close to the 0.10 significance level ($P\text{-value} = 0.1359$). Miscellaneous expenses per interview for falsifiers were nearly twice as

much as expenses per interview for nonfalsifiers. In addition, it is interesting to note the value differences between some of the factors, although these differences are not statistically significant. For example, the mean number of cases for falsifiers was 92.78 compared to 83.74 for nonfalsifiers. Less experienced FIs with a larger workload may turn to falsification to meet project requirements and complete a full assignment. The hours per interview were very similar between falsifiers and nonfalsifiers but the miles per interview were less for falsifiers than nonfalsifiers. One possible interpretation is that falsifiers may not actually travel to the household to complete the work so their reimbursable mileage is lower than FIs who did.

Table 1. Difference of Means Between Falsifiers and Nonfalsifiers for Less Experienced Interviewers

Variable	Falsifiers (N=65)	Nonfalsifiers (N=47)	P-Value	Standard Error
Workload (Cases)	92.78	83.74	0.2527	7.8623
Field Interviewer Screening Response Rate (%)	82.79	85.69	0.2232	0.0254
Field Interviewer Interview Response Rate (%)	70.65	70.10	0.8481	0.0300
Miscellaneous Expenses/Interview (\$)	Below Average	Below Average	0.1359	1.8844
Hours/Interview (Hours)	Average	Average	0.6147	1.0877
Miles/Interview	85.33	95.38	0.5220	14.7441
Screening Error Rate	0.19	0.13	0.0732	0.0306
Interview Error Rate	0.43	0.35	0.4728	0.1035

Difference of Means Tests Between Falsifiers and Nonfalsifiers Among More Experienced Interviewers:

Difference of means tests were also used to compare the workload, response rates, cost per interview, and error rates of falsifiers with those of nonfalsifiers among more experienced FIs to determine if any of those factors were associated with falsification among more experienced FIs. There were significant (P-value<0.05) differences in the workload, hours per interview, and error rates between FIs who falsified and those who did not. Table 2 displays the results. Among experienced FIs, those who falsified had a smaller workload than FIs who did not falsify (P-value=0.0156). Falsifiers also had a higher hours per interview (P-value=0.0252) than nonfalsifiers. Among the more experienced FIs, those who falsified had a smaller workload than those who did not falsify but higher hours per interview. FIs who falsified may have become frustrated with the amount of time

spent attempting to complete screenings or interviews among a smaller number of cases. This frustration could have led to the decision to falsify data in order to complete a time-consuming case. In regards to error rates, FIs who falsified had higher screening and interview error rates (P-value=0.0308 and 0.0144 respectively). A consequence of falsifying data may be more data quality errors. Therefore, it is not surprising that FIs who falsified have higher error rates than those who did not.

An interesting point is that among the less experienced FIs, there was only a significant (at the 0.10 level) difference in the screening error rate between falsifiers and nonfalsifiers but there were significant (at the 0.05 level) differences in both screening and interview error rates when comparing falsifiers and nonfalsifiers among more experienced FIs. When comparing the error rates of less experienced FIs to more experienced FIs, the error rates are higher for less experienced FIs. Newer FIs may be more prone to data quality errors given their lack of experience. These errors could be unintentional and the result of not paying attention to project details. However, when more experienced FIs have higher error rates, this may be a better indication of falsification among that group.

Table 2. Difference of Means Between Falsifiers and Nonfalsifiers for More Experienced Interviewers

Variable	Falsifiers (N=24)	Nonfalsifiers (N=38)	P-Value	Standard Error
Workload (Cases)	127.8	178.9	0.0156	23.3675
Field Interviewer Screening Response Rate (%)	87.74	89.36	0.3432	0.0171
Field Interviewer Interview Response Rate (%)	74.53	79.04	0.1412	0.0275
Miscellaneous Expenses/Interview (\$)	Above Average	Above Average	0.3501	5.8247
Hours/Interview (Hours)	Above Average	Below Average	0.0252	0.8860
Miles/Interview	90.69	81.89	0.4863	12.5671
Screening Error Rate	0.11	0.07	0.0308	0.0170
Interview Error Rate	0.27	0.14	0.0144	0.0440

CONCLUSIONS

When comparing FIs who falsified to those who did not among less experienced FIs, there were differences in the miscellaneous expenses per interview and the screening error rate. Both factors were higher for falsifiers than nonfalsifiers. This same comparison among more experienced FIs resulted in significant differences in workload, hours per interview, and error rates between FIs who falsified and those who did not. FIs with more experience who falsified had a smaller workload, higher hours per interview, and higher error rates than experienced FIs who did not falsify.

FUTURE CONSIDERATIONS

While this research provides insight into the differences among field verified FIs on less studied factors, there is more that can be done to understand the possible associations between field interviewer costs, error rates and falsification on the NSDUH. It may be beneficial in the future to examine error rates for FIs for whom a higher percentage of work was selected for verification and compare the error rates among that group of FIs who were found to be falsifying and FIs who were not. Another possible investigation into patterns on NSDUH could be to further divide the less experienced FIs by weeks on the project. In the current analysis, there were significant differences in the expenses per interview and screening error rate between falsifiers and nonfalsifiers among FIs working less than 52 weeks. Would the results be different if you looked at falsifiers and nonfalsifiers among FIs having worked less than 26 weeks? Would significant differences in additional factors come to light?

Future considerations also include expanding the investigations to other projects. Would there be different results depending on the project? Would there be significant differences in workload, cost per interview, response rates, and error rates among less experienced FIs on other projects? Comparing the characteristics of the entire staff of field interviewers on NSDUH to another project may demonstrate that NSDUH has unique circumstances. Comparing how field verifications are conducted on different projects as well as the criteria used to determine whether one is warranted are other investigations to consider. There are ample directions for future research which may offer a richer and more detailed picture of factors associated with falsification.

Given the relationship between falsification and FI experience in previous research, understanding the reasons that less experienced field interviewers turn to falsification to complete their work would allow field supervisors to be proactive in curbing falsification activities before they begin. Are they overwhelmed by the job?

Do they not feel supported by their field supervisor? Are they not prepared for the refusals they will face in the field? Research into the reasons for falsification could allow project management to be even more proactive in their approach. Additional coaching, training and support of field interviewers could lessen the likelihood they turn to falsification.

REFERENCES

- Bredl, S., Winker, P., and Kotschau, K (2008). A Statistical Approach to Detect Cheating Interviewers. Center for International Development Research and Environment Research of the Justus-Liebig-University of Giessen.
- Judge, G. and Schechter, L (2009). Detecting Problems in Survey Data Using Benford's Law. *Journal of Human Resources* 44(1): 1-24.
- Murphy, J., Baxter, R., Eyerman, J., Cunningham, D., and Kennet, J (2004). A System for Detecting Interviewer Falsification. Proceedings of the American Association for Public Opinion Research, Section on Survey Research Methods.
- Porras, J., and English, N. (2004). Data-Driven Approaches to Identifying Interviewer Data Falsification: The Case of Health Surveys. Proceedings of the American Statistical Association, Section on Survey Research Methods.
- Schrapler, J. and Wagner, G (2003). Identification, Characteristics and Impact of Faked Interviews in Surveys – An Analysis by Means of Genuine Fakes in the Raw Data of SOEP. IZA Discussion Paper Series, 969.
- Schreiner, I., Pennie, K., and Newbrough, J (1988). Interviewer Falsification in Census Bureau Surveys. Proceedings of the American Statistical Association, Section on Survey Research Methods, pp. 491-496.
- Swanson, D., Cho, M.J., & Eltinge, J. (2003). Detecting Possibly Fraudulent or Error-Prone Survey Data Using Benford's Law. Proceedings of the American Statistical Association, Section on Survey Research Methods.