INTERVIEWER FALSIFICATION IN CENSUS BUREAU SURVEYS

Irwin Schreiner, Karen Pennie, Jennifer Newbrough, U.S. Census Bureau 1/
Irwin Schreiner, U.S. Bureau of the Census, Washington, D.C. 20233

1. BACKGROUND

Over the years it has been the policy of the Census Bureau to include in continuing and one-time surveys, a reinterview program to control the quality of interviewing. For continuing surveys the reinterview program is designed to provide feedback to the interviewers regarding their performance. For the most part the level of performance of the interviewing staff meets or exceeds the acceptable standard. For those who fail to meet this standard some form of remedial training is provided, unless they are identified as falsifying data. For most one-time surveys, on the other hand, there is generally very little time for providing any feedback to interviewers. For these surveys, the use of the reinterview as a control on interviewers is strictly to identify those who are falsifying data.

Initially, the impetus for gathering information on interviewer falsification came from the need to design an optimal reinterview sampling strategy for one-time surveys. Due to budget constraints, it was generally not possible to reinterview every interviewer and at the same time select enough cases to achieve a high probability of detecting minimal falsification. Sometimes it was suggested only new interviewers be checked. The argument was made that new interviewers were the ones most likely to falsify data and the more experienced ones could be trusted. Of course, this was merely an intuitive belief, since there was no empirical data to back it up.

Following the 1980 Decennial Census a second and stronger impetus came from the Reinterview Work Group (RWG). This group had the task of redesigning the current surveys' reinterview program. However, it too was hampered in the effort to design an optimal sampling strategy for interviewer control. This was primarily because of the lack of factual information concerning the nature of interviewer falsification. At that time there was no data available on falsification other than anecdotes reported by field supervisors and these experiences were not systematically captured.

One particular piece of knowledge that would have been quite valuable was whether or not interviewer falsification was clustered. If, in fact, it was, then resources could best be used by spreading the reinterview sample over as many clusters as feasible, given cost constraints, rather than the existing procedure of selecting entire clusters.

Furthermore, if interviewers who falsify data tend to do so infrequently, but on a large proportion of their assignment, then the best strategy is to reinterview a small amount of their assignment fairly often. If, on the other hand, interviewers who falsify data, tend to do so on a small proportion of their assignment quite often, then the best strategy is to reinterview as much of their assignment as possible less frequently.

These and other questions regarding interviewer falsification finally led to the implementation in August 1982 of the Interviewer Falsification Study.

2. INTRODUCTION

This paper covers the results from the first five years of the study; that is, from September 1982 through August 1987. The results were tabulated from an Interviewer Follow-up Form (IFF) that the twelve Census Bureau regional offices were instructed to fill each time a case of falsification was suspected or confirmed for any survey.

Falsification is defined as occurring whenever the interviewer knowingly deviates from current interviewing procedures to avoid interviewing, classifying, and/or listing units. Falsification might also include the acceptance of proxy information when self-response is required and the unauthorized use of the telephone when a personal visit is required. These latter instances are considered falsification when the interviewer knowingly deviates from the current procedure and attempts to conceal this fact.

With the Interviewer Falsification Study having accumulated data for a five year period, certain patterns of falsification have emerged. These patterns answer some of the questions raised concerning the effectiveness of the reinterview program. They also give insights into the development of current and future reinterview sampling strategies. Reassuringly, from what has been learned at the Census Bureau, the incidence of falsification in surveys is quite low. Thus, while certain patterns have developed, it will take more data to determine 1) whether these patterns are significant and 2) if others exist. As such, this research is expected to continue for at least another several years.

3. SUMMARY OF RESULTS

3.1 General Results

3.1.1 Number of Interviewer Follow-up Forms (IFF) Received

Over the first five years of the study, 246 IFF's were received, of which 205 (83 percent) involved instances of confirmed falsification. The remaining 41 forms represent cases where falsification could not be confirmed or it was determined falsification did not occur. They are not included in this analysis.

3.1.2 Method of Detection

The method of detection and the corresponding number of detected interviewers for the 205 cases of confirmed falsification are shown in Table 1. Most of these, 151 (74 percent), were detected through reinterview. For the remaining 54 cases, the method of detection was mostly by chance, with 36 cases detected because "something out of the ordinary occurred." Seventeen of these surfaced when when an interviewer or supervisory field representative (SFR) took over the assignment of an interviewer who became ill, left the Census Bureau, etc. The surveys in which these cases were discovered were panel surveys where the same households are interviewed at specified intervals.

Table 1. Method of Detection

Method	Number
Reinterview	151
Edit of forms	12
Unusual noninterview rates	5
Unusual production ratios	1
Something out of the ordinary	36
Total	205

3.1.3 Falsification by Survey

The 205 cases of confirmed falsification were distributed across 15 surveys conducted by the Census Bureau (see Table 2). Five of these surveys accounted for 166 (81 percent) of the cases. These were the Current Population Survey (CPS) with 71 cases, the National Crime Survey (NCS) with 33 cases, the New York City Housing Vacancy Survey (NYC-HVS) with 27 cases, the American Housing Survey Metropolitan Sample (AHS-MS) with 20 cases and the Survey of Income and Program Participation (SIPP) with 15 cases. The fact that CPS has the most cases is not at all surprising since it has more interviewers than any other survey.

Through the use of the reinterview, rates of falsification were computed for CPS, NCS, and NYC-HVS. These rates provide insight into the extent of falsification here at the Census Bureau. For CPS and NCS estimated monthly rates of falsification were computed. These rates were obtained by taking the number of falsifiers detected through the regular reinterview and dividing this by 60 (months of the study) to get an average number detected per month. The result was then multiplied by the inverse of the proportion of interviewers in reinterview each month and finally dividing this number by the average number of interviewers working on the survey each month. For example, in NCS the rate was computed as follows:

$$\frac{\text{18 detected through regular reinterview}}{\text{60 months of the study}} \times \frac{\text{6 inverse of proportion in reinterview}}{\text{473 aver/mo working on NCS}} = 0.4\%$$

For CPS the result was also 0.4 percent. These rates are conservative in that they do not take into consideration the probability of detection. When an interviewer is selected for reinterview in NCS one-third of the assignment is reinterviewed. The same applies for CPS, except that beginning with January 1987 this was reduced to only one-fourth of the assignment.

For the NYC-HVS, the rate for 1984 and 1987 combined was computed by dividing the number of interviewers detected by the number who worked on this survey. This was because the NYC-HVS had a continuous reinterview program in which cases were selected from each transmittal for each interviewer. The resulting rate was 6.5 percent.

These rates, of course, are at the interviewer level. For those interviewers caught falsifying data, the actual amount of work falsified is covered in section 3.1.5.

It should be pointed out that CPS, NCS, and SIPP have regular interviewing staffs that conduct interviews each month. The AHS-MS is set up so that interviewing within a specified metropolitan area is conducted once every four years. It has monthly interviewing assignments extending over a period of six to eight months. The NYC-HVS is conducted once every three years and interviewing is completed within a period of three months or less. Thus, due to the nature of these latter two surveys, the interviewing staff is made up almost exclusively of new hires who are released at the conclusion of the survey.

Table 2. Falsification by Survey

Survey	Number
Current Population Survey (CPS)	71
National Crime Survey (NCS)	33
New York City Housing Vacancy Survey (NYC-HVS)	27
American Housing Survey Metropolitan Sample (AHS-MS)	20
Survey of Income and Program Participation (SIPP)	15
Others (10 surveys)	39
Total	205

3.1.4 Type of Falsification

Falsification of data can take on many forms ranging from the most severe form of actually fabricating interviews to certain deviations from interviewing procedure such as the violation of personal visit rules. Basically, if interviewer errors are not the result of being uninformed, but rather are the result of being dishonest, it is considered falsification. Of the 205 interviewers caught falsifying data, 162 (79 percent) were determined to have fabricated interviews.

Table 3 shows the type of falsification, cross tabulated by survey. Since there are times when an interviewer may be caught deviating from current interviewing procedures in more than one way, the total across all types of falsification exceeds the total number of falsifiers. Besides the fact most falsifiers fabricate interviews, there are two other interesting outcomes shown in this table. First of all, the second highest type of falsification was deliberately misclassifying units as vacant when they were occupied. Thirty-eight interviewers were discovered doing this. This is not surprising when one understands that interviewers are evaluated on their type A noninterview rate. Type A's are occupied households for which an interview cannot be obtained. Second, NCS, which requires self response for each household member 12 years of age and older, has a particular problem with interviewers recording self response when they interviewed a proxy. Twenty-four interviewers were caught doing this.

3.1.5 Percent of Assignment Falsified

Over the first three years of the study, the percent of the assignment falsified was determined strictly through estimates made by the field personnel responsible for reporting it. The question on the IFF pertaining to the percent of the assignment falsified asked whether the entire assignment was falsified or a portion of the assignment was falsified. When only a portion of the assignment was involved, an estimate of the percent falsified was requested.

After the third year of the study the IFF was revised to include a table to be filled with the number of units in the assigned clusters and the number of units falsified. By supplying the actual assignment and the number of units falsified, it was felt that better estimates of the percent of the assignment falsified could be obtained. However, there was still one remaining weakness in the estimates of the percent of the assignment falsified. Not all of the units in the assignment of each falsifier were always examined.

As a result, after the fifth year of the study the table was further revised to include only those households actually examined. In the meantime, in an effort to obtain some estimate of the average percent of the assignment falsified, the results over the last two years of the study from the regular reinterview for two surveys, CPS and NCS, were used. Sixteen falsifiers were involved, 13 from CPS and 3 from NCS. Their combined estimate was 31.4 percent (median = 30.5 percent). The standard deviation on this estimate of the average was 4.2 percent.

3.1.6 Length of Service

One area of particular interest in our examination of interviewer falsification was the nature of its relationship to length of service. This examination was restricted to CPS and NCS because of 1) the ongoing nature of these surveys, 2) the availability of interviewer data, and 3) the method of selecting interviewers for reinterview.

Table 3. Falsification by Type and Survey

Survey	Fabricated Interviews	Violation of Rules: Self Response	Violation of Rules: Personal Visit	Occupied Classified as Vacant	Interviewed Ineligible Proxy	Other
AHS-MS	17	NA	2	6	0	0
CPS	58	NA	2	18	8	2
NCS	12	24	9	7	3	1
NYC-HVS	27	NA	0	3	1	0
SIPP	15	NA	2	0	2	0
Others (10)	33	1	3	4	2	5
TOTAL	162	25	18	38	16	8

The overall average length of service of all interviewers for CPS and NCS combined was 6.22 years. This average was computed using the number of interviewers on board and their average length of service for each survey as of October 1, 1985, 1986, and 1987. This staff comprises the total universe of interviewers from which the reinterview was selected. In the study there were 16 interviewers who were detected through the regular reinterview as falsifying data at some point during the period of October 1985 through August 1987 in either CPS or NCS. Their combined average length of service was 1.72 years (median = 1.67). The standard deviation on this estimate of the average was 0.26 years. This means that the average length of service of falsifiers is significantly lower than for all interviewers.

3.2 Detailed Results

It was pointed out at the beginning of this paper that knowledge of certain information regarding falsification would be valuable in the design of reinterview programs. This was the falsification patterns of new versus experienced interviewers and whether or not falsification was clustered. This section summarizes the findings in these two areas.

3.2.1 New Versus Experienced Interviewers

Cross tabulations were done by length of service with several other variables. It was hypothesized that differences existed between the falsification patterns of new interviewers (less than one year of service) and those of experienced interviewers (a year or more of service). This analysis was restricted to two surveys, CPS and NCS, since the other surveys had either too few cases or almost all the interviewers had less than one year of service.

The comparisons which follow use the results from all the falsifiers for these two surveys regardless of how they were detected. Since very few of the newer interviewers were caught falsifying data through the regular reinterview, we were left with very small sample sizes from this group. This made it unfeasible to do any statistical testing. Still, we feel the differences which emerged are meaningful ones.

3.2.1.1 Falsification in Continuing Households Only

In CPS only 38 percent (5 out of 13) of the new interviewers fabricated interviews in continuing households only as compared to 73 percent (33 out of 45) of the experienced interviewers. This apparent difference suggests the more sophisticated approach taken by the experienced interviewers. CPS is a panel type survey where the same housing unit is interviewed monthly. The data gathered can often be correctly "imputed" from a past month to the current month. Therefore, there is less risk of an interviewer being caught falsifying data in a household that has previously been interviewed (continuing) as opposed to a household in sample for the first time (new).

For NCS this same pattern did not show up. This is not surprising, since in NCS the interviews for the same housing unit are six months apart and the data gathered concerns incidents of crime occurring over the previous six months.

3.2.1.2 Falsification in Units Requiring Personal Visits

In NCS none of the 12 new interviewers caught falsifying data did so in only those households requiring personal visits as compared to 33 percent (7 out of 21) of the experienced interviewers. This difference suggests the more selective approach to falsification taken by experienced interviewers.

In CPS the overiding issue is the tendency of experienced interviewers to fabricate interviews in continuing households only. Since new households require personal visits and continuing households can for the most part be done by telephone, a similar pattern would not be expected to exist.

3.2.1.3 Type of Falsification

In NCS certain differences were revealed between new and experienced interviewers by type of falsification. These are summarized in Table 4.

Table 4. Percentage of NCS Falsifiers by Type of Falsification

Type	Years of Service: Less Than One	Years of Service: One Year or More
Fabricating Interviews	58.3	23.8
Violation of Rules:		
Self-Response	66.7	76.2
Personal Visit	50.0	14.3
Total Falsifiers	12	21

Since an interviewer may commit more than one type of falsification, these percentages by themselves do not tell the whole story. The more interesting statistic is the average length of service. For the 12 interviewers caught fabricating interviews it was 1.1 years

(median = 0.8 years). For the nine interviewers caught violating personal visit rules it was 1.7 years (median = 0.6 years). For the 24 interviewers caught violating the self-response rules it was 4.8 years (median = 2.8 years), and for the ten interviewers whose only type of falsification was violation of the self-response rules, the average length of service jumps to 9.3 years (median = 7.9 years).

3.2.2 Clustering

During the first three years of the study the item pertaining to clustering merely asked if the interviewer usually falsified entire clusters or consistently the same number of units in some or all of the clusters. After the third year of the study this item was replaced with a table, mentioned earlier, which provided information regarding the number of units falsified in a particular cluster.

The analysis relating to clustering was restricted to two surveys, CPS and NCS. This was because AHS-MS is made up primarily of individual units and SIPP comprises mostly clusters of size two. CPS and NCS have compact clusters, or segments, containing on the average four adjacent housing units.

Initially, a fairly crude method of measuring clustering was used. This method was the only one which could be applied to the first three years of the study and was extended over the last two years to maintain consistency. Under this method a high amount of clustering was equated with falsifying entire segments and a moderate amount of clustering was equated with more than one unit in all falsified segments. Anything less was considered low. An attempt was made to recode responses from the "other" category into one of the above categories. If the response could not be recoded, such as when only one unit in the entire assignment was falsified, it was classified as no clustering. Table 5 shows most interviewers caught falsifying exhibited low clustering or none at all. This applied to both CPS and NCS.

Table 5. Percent of Falsifiers by Amount of Clustering

Degree of Clustering	CPS	NCS
High	12.6	12.1
Moderate	8.5	3.0
Low	52.1	48.5
None	19.7	33.3
NA	1.4	0.0
Unknown	5.6	3.0
Total Falsifiers	71	33

For the two most recent years of the study, an intraclass correlation coefficient was calculated to measure the clustering of falsified cases. A high intraclass correlation (close to 1.0) means that interviewers falsified a large proportion of cases in each segment. If a small proportion was falsified, the intraclass correlation will be low positive or even negative.

The intraclass correlation coefficient was computed using the reinterview sample segments in the assignments of those falsifiers discovered during the regular reinterview. To qualify a segment had to contain at least two housing units. In CPS there were 47 segments used in the analysis and eight segments from NCS. This involved a total of 16 interviewers. The result was 0.27. This coeffiecient is small thus indicating most falsification exhibits low clustering. The 90 percent confidence interval on this estimate ranges from a very low level of 0.11 to a moderate upper level of 0.42.

4. LIMITATIONS OF THE DATA

The major limitation of this study has been the small sample sizes available for analysis. This is especially true in the comparisons of the new and experienced interviewers and the computation of the intraclass correlation coefficient. For the most part in Section 3.2 we had to use not only the results from those interviewers caught falsifying data through the regular reinterview, but also from those caught through other means. This precluded us from doing any statistical testing. In the analysis of the average length of service the comparison was restricted to the results from the last two years of the study. This was because of the unavailability of complete data on the nature of the full interviewing staff prior to October 1, 1985.

A second limitation of this study was in the initial design of the IFF. The original form was not effective in obtaining accurate estimates of the percentage of the assignment falsified, nor did it obtain data on clustering in the most appropriate manner. Even the major revisions to the IFF after the third year of the study still left some minor problems with the data gathered. The most recent revisions should take care of this.

A third and ongoing limitation of the study has been in the quality of the data collected. We must rely on the supervisors in the field to fill out an IFF each time a case of falsification is discovered. If they do not, this data is lost. Even when the IFF's are sent in there have been numerous instances in which they have had to be returned because certain information was either missing or inconsistent. Finally, even when the form has been completely

filled, we are not always sure the supervisors are recording what has been requested. We suspect in some cases the instructions have been misunderstood. We plan to examine this problem further.

Fourth and finally, we will always be limited by the fact that not all falsifiers are caught. As such, the results only represent the population of falsifiers that are detected through the reinterview and various other informal means.

In spite of all these limitations we feel we have been successful in creating a database on falsification. The results obtained thus far, while they contain a number of flaws, give us some indication of the nature and degree of interviewer falsification. In addition, they highlight potential problems we can examine more closely.

5. CONCLUSIONS

The results from the study provide fairly substantial evidence that the shorter the length of service the more likely it is an interviewer will falsify data. Since this was based upon data from the two most recent years, some further investigation is in order.

At the same time, certain differences in the falsification patterns of new and experienced interviewers were revealed. They indicate a certain selectivity on the part of the experienced interviewers directly related to the nature of CPS and NCS. This is demonstrated in the former by the increased likelihood of falsification in continuing households only and in the latter by the preponderance of falsification in only personal visit households. In NCS this selectivity also extends to the type of falsification, where it was found that all of the interviewers whose only type of falsification was violation of the self-response rules were experienced.

Finally, the results to date have indicated that when an interviewer falsified a portion of the assignment there was little or no clustering. This area in particular will require careful examination over the next year or two, since the intraclass correlation coefficient was computed on a rather small sample.

With regard to the reinterview sample design the above results indicate that for the newer interviewers it may be useful to reinterview some of their work more frequently. For the experienced interviewers it appears that concentrating on those aspects of a particular survey which are more likely to lead to falsification will provide the most payoff. Because the level of clustering appears to be low or none at all, this gives good reason to believe the current reinterview sample design is appropriate.

REFERENCES

Jones, Charles D. (1981), "Developing a Profile of a Curbstoner," unpublished memorandum to Lawrence T. Love, Washington, D.C.: U.S. Bureau of the Census

Love, Lawrence T. (1982), "Curbstoning Study," unpublished memorandum to All Regional Directors, Washington, D.C.: U.S. Bureau of the Census

Poisson, Michael J. (1987), "Results Memorandum for the 1987 New York City Housing and Vacancy Survey," unpublished memorandum to Sheila H. Goehringer, New York, N.Y.: U.S. Bureau of the Census

Reinterview Redesign Report 1: Optimal Sampling Strategy for Interviewer Control (1982), unpublished report, Washington, D.C.: U.S. Bureau of the Census

Schreiner, Irv (1980), "Proposal for Developing a Profile of a Curbstoner," unpublished memorandum to Robert T. O'Reagan, Washington, D.C.: U.S. Bureau of the Census

Schreiner, Irv and Newbrough, Jennifer (1987), "Falsification Study (September 1982 through August 1987)," unpublished report, Washington, D.C.: U.S. Bureau of the Census

Smith, Robert T. (1987), "Clustering of Falsified Data Cases," unpublished memorandum for The Record, Washington D.C.: U.S. Bureau of the Census

The authors wish to acknowledge the contributions of Mary Ellen Beach, Richard Blass, Michael Haas, Alice Harrington, Sadie Jones, Patricia Rogers, and Robert Smith

1/ This paper reports the general results of research undertaken by Census Bureau staff. The views expressed are attributable to the authors and do not necessarily reflect those of the Census Bureau.