USING DATE AND TIME STAMPS TO DETECT INTERVIEWER FALSIFICATION

John M. Bushery, Jennifer W. Reichert, Keith A. Albright, John C. Rossiter U.S. Census Bureau John M. Bushery, U.S. Census Bureau, DSMD, Washington, D.C. 20233 john.m.bushery@ccmail.census.gov

Key Words: reinterview, CAI, instrument, quality control (QC)

Falsification and Interviewer Quality Control

Interviewer data falsification is an important concern at the U.S. Census Bureau. Fortunately, the Census Bureau's emphasis on interviewer quality control keeps falsification a rare event. Only about one interviewer in 100 is caught falsifying data in our ongoing surveys (Wetzel, 1993). Over the years, the Census Bureau has developed a toolkit for interviewer quality control. The availability of date and time stamps from Computer Assisted Interviewing (CAI) provides another possible tool to detect falsification. This paper describes research conducted in the National Health Interview Survey (NHIS) to determine whether patterns in date and time stamps correlate with falsification of interviews.

Reinterview Programs

The Census Bureau's main tool for controlling data falsification is its quality control random reinterview. This reinterview program checks between two and 10.0 percent of the cases in each survey to assure the interviews were actually conducted. In this reinterview, senior interviewers or supervisors use the telephone to check that the interviewers actually conducted the interviews. They check by personal visit if necessary.

The Census Bureau has used reinterview programs to detect and deter interviewer falsification since the 1950s. Interviewers are selected at random and a sample of their work is checked. In 1982 the Census Bureau developed a falsification database to guide improvements to the reinterview program (Schreiner, Pennie, and Newbrough, 1988). The database allows the Census Bureau to model interviewer behavior making the reinterview more effective and efficient (Biemer and Stokes, 1989). The most recent improvement is the "focused reinterview." The fast turnaround of interviews under CAI allows near real-time statistical analysis of survey data to identify problems in the interview process. The Census Bureau began the focused reinterview program in 1997 to improve detection of interviewer falsification (Hood and Bushery, 1997). CAI can also capture "date and time stamps" during interviews. This research aims to determine whether the focused reinterview program can be extended to use date and time stamps to detect falsification.

Focused Reinterview

The focused reinterview supplements the random reinterview program. This system analyzes interview data to identify "questionable" interviewers. These interviewers' assignments contain a relatively high proportion of cases "at risk" for falsification. These atrisk cases are sent back to the field for a special reinterview check -- the "focused reinterview." The cornerstone of the focused reinterview is a falsification model with two main premises:

- Minimal effort -- interviewers who falsify data will try to keep it simple and fabricate a minimum of data.
- They also will try to make it more difficult to check falsified cases.

We used a "content-based" focused reinterview experimentally in the 1995 American Travel Survey, the 1996 Survey of Fishing, Hunting, and Wildlife Associated Activity, and the 1997 NHIS. We have since used it on a production basis in the 1998 NHIS and the 1998 Survey of Program Dynamics (SPD). The contentbased focused reinterview identifies interviewers with unusual work with respect to the data collected, or not collected, in the interview (e.g., no phone number rate, noninterview rate, screen-out rate, etc.).

Because falsification is so rare, statistical testing reveals no significant differences between detection rates for the random and focused reinterview so far. However, early results suggest that the focused reinterview may detect falsification more efficiently than the random reinterview (Hood and Bushery, 1997). For example, the focused reinterview checked 239 NHIS interviewer assignments and found four instances of falsified data

This paper reports the results of research and analysis undertaken by Census Bureau staff. It has undergone a more limited review than official Census Bureau Publications. This report is released to inform interested parties of research and to encourage discussion.

(2.0 percent). The NHIS random reinterview confirmed falsification for only one of the 708 interviewer assignments checked (0.1 percent). Similarly, the SPD focused reinterview found one falsifier among the eight interviewers checked (13.0 percent), while the SPD random reinterview found two falsifiers among the 262 interviewers checked (0.8 percent).

Date and Time Stamps

The CAI instrument can store the date and time of any activity in the instrument. The NHIS instrument has been written to capture the date a case was last accessed. We assume this "date stamp" indicates when the interviewer completed the interview.

The NHIS instrument contains time stamps in several locations. "Time stamps" are actually cumulative time measurements. They measure the elapsed time for the entire interview and for selected sections of the interview. The NHIS time stamps cover these sections:

- household section
- sample adult section
- sample child section
- child immunization section

Date and time stamps may increase the sensitivity of the focused reinterview screening. The "minimal effort" aspect of the falsification model suggests that interviewers might complete all fabricated interviews on the same day. An interviewer who completes more cases on the same day than is reasonable may be fabricating interviews. We would select these cases for a focused reinterview check to assure that the interviewer actually conducted all the interviews.

Interviewers might feel tempted to fabricate interviews if much of their workload still remains unfinished in the last days of the assignment period. An unusually large number of cases completed near the end of the assignment period might mean the interviewer falsified these cases.

Time stamps can reveal extremely short interviews. We hypothesize that fabricating an interview would be faster than a comparable real interview. An interviewer with significantly more short interviews than the norm would also receive a focused reinterview check.

Evaluation Methods

We used statistical process control (SPC) charts (Grant and Leavenworth, 1988) to identify interviewers who completed more cases than expected for specified time periods. We compared weekly interviewer assignments against a historical average for the first 17 weeks of data collected in 1999.

These statistical methods screen interviewers to determine whether a focused reinterview check is warranted. To be useful this statistical screening should meet three criteria:

- It identifies "outlier" interviewers who deviate significantly from the norm.
- It doesn't identify too many outlier interviewers for the reinterview check. (The process is not in statistical control if the analysis identifies too many outliers.)
- The "date stamp" reinterview check returns a higher proportion of falsification than the random reinterview.

This paper discusses the first two criteria. Determining whether our statistical screening meets the third criterion will require considerable time. We plan to conduct focused reinterview checks to determine whether the outlier interviewers' cases were falsified.

First, we will compare the list of interviewers flagged by the date and time stamp analysis to the list of interviewers flagged by the content-based focused reinterview analysis (Hood and Bushery, 1997). The degree of overlap between the lists can shed some insight into the level of success of the date and time stamp methods. If the date and time stamp analysis flags only interviewers identified as "questionable" by the contentbased focused reinterview analysis, then there are no gains from analyzing the date and time stamps.

However, if we identify interviewers who weren't previously flagged in the focused reinterview, we will check those interviewers in the focused reinterview. The results of the focused reinterview for these interviewers will ultimately show whether date and time stamps meet the third criterion.

Date Stamps

We look for suspicious patterns in the date stamps associated with each interviewer's assignment. The characteristics of interest are:

- An extreme number of cases completed in one day.
- An extreme number of cases completed during any three days of the assignment period.
- An extreme number of cases completed at the end of the assignment period.

All three of these analyses include only full and partial interviews and ineligible cases in the count of "completed" cases. Ineligible cases are cases where there are no eligible respondents to interview (e.g., vacant households, demolished homes, etc.). Noninterviews (e.g., refusals, no one home, etc.) don't count in the number of completed cases.

Extreme Number of Cases Completed in One Day

The average NHIS interview lasts over an hour. Therefore, it is reasonable to expect that an interviewer can complete a limited number of interviews in a single day. If an interviewer completes far more cases in a day than the average for the regional office, that interviewer may be falsifying the data.

An interviewer probably can complete a fabricated interview much faster than a real interview, allowing the interviewer to complete more cases in a day. A real interview requires time to contact the respondent (the NHIS is primarily a personal visit survey), and during the interview, distractions are likely, which may cause the interview to take longer (e.g., children, pets, phone calls, etc.).

To identify interviewers who complete an unrealistic number of cases in a day, we computed the maximum number of cases completed per day for each interviewer and the average maximum over all interviewers in the regional office. We flagged as an outlier any interviewer whose maximum was more than three standard deviations above the regional office's average maximum.

Extreme Number of Cases Completed During Any Three Days

This analysis used SPC charts (\bar{X} , R) (Grant and Leavenworth, 1988) to identify interviewers who complete more cases than the RO average in their "top three" days. If an interviewer's top three average was more than three standard deviations above the regional office's top three average, we flagged that interviewer as an outlier. Here \bar{X} represents the mean number of cases completed in the interviewer's top three days and R represents the range for the top three days.

Extreme Number of Cases Completed at the End of the Assignment Period

We used SPC P-charts with 3-sigma control limits (Grant and Leavenworth, 1988) to identify interviewers who returned an extreme proportion of ineligible cases (vacant, not a housing unit, screened out as non-minority) at the end of the interview period. Because ineligible cases do not count against an interviewer's response rate, an interviewer caught in a crunch at the end of the assignment period might be tempted to classify eligible cases as ineligible. We defined "end of the assignment period" as the last officially scheduled day of the assignment period. However, the assignment period usually was extended by a day, so the "end of the assignment period" often consisted of two days.

We performed analyses using two different proportions. All proportions used the interviewer's full workload as the base.

- Interviews completed at the end of the assignment period as a proportion of all cases in the assignment.
- Ineligible cases completed at the end of the assignment period as a proportion of all cases in the assignment.

Time Stamps

This paper will not address the time stamp data. However, the time stamps may enable us to detect performance problems as well as falsification. We plan to evaluate how much time interviewers spend conducting each interview. This information may provide clues as to whether or not an interviewer actually visited the household and conducted the interview.

The NHIS interview covers complicated issues in detail, such as recent doctor visits, injuries, and health insurance. These interviews generally require an hour or more to complete. As we mentioned earlier, fabricated interviews would take less time to complete. Interviewers who spend significantly less time than average conducting interviews may be falsifiers.

On the other hand, interviewers who spend too much time conducting interviews may require further training. Interviews that take longer than necessary increase respondent burden and costs. Field supervisors could observe interviewers whose interviews consistently take longer than average to determine whether that interviewer needs additional training or coaching.

Limitations

Interviewer behavior and laptop computer malfunctions can affect the reliability of date and time stamp data.

Our research assumes that the date stamp reflects the date that the interviewer actually completed the interview. However, because they record the date the case was last accessed, the date stamps are not always 100 percent accurate. If an interviewer reenters a completed case, say simply to view the notes for that interview, the instrument will reset the date stamp. For example, if an interviewer completes a case on Monday, but reenters it on Tuesday, the date stamp will reflect the Tuesday access. Fortunately, interviewers are required to transmit their

completed cases every night, so this problem should be negligible.

The time stamp data probably are more subject to error. Software or hardware malfunctions during the interview may affect the accuracy of time stamp data. Because the time stamps are a cumulative measure of time, the stamps require clear start and end times. If the interviewer does not exit the instrument properly (for example, reboots the computer while still in the instrument), the instrument will not record end times for the current interview section or for the entire interview. These types of malfunctions affect both the time stamp for the entire interview and the time stamp for the section of the interview where the malfunction occurred.

When the interviewer reenters and completes a case affected by a laptop malfunction, the time stamp for the entire interview will reflect the second entrance into the case and a shorter interview than actually occurred.

Such malfunctions do not affect time stamps for sections of the interview completed prior to the malfunction. So the sum of time stamps for the individual sections can exceed the time stamp for the entire interview.

We consider impossibly short interviews and unreasonably long interviews to be ineligible for this analysis. We may eliminate interviews shorter than five minutes or longer than four hours from the analysis. We also will explore using the sum of section time stamps to replace the overall time stamp when the section sum is larger.

Results

Table 1 summarizes the results of the date stamp analysis. As we identify outlier interviewers, we will send the appropriate cases to the field for the focused reinterview check.

The analysis of the "top three days" proved disappointing. It failed to satisfy the second criterion -it identified too many outliers -- about 8 per week. That is too many to send out for the "date stamp" focused reinterview check. The random reinterview selected only 12 interviewers per week in the same time period. The focused reinterview currently identifies about two interviewers per week.

Identification of too many outliers is an indication that the process is not in statistical control for the "top three days" data. There is too much variability in the data for SPC methods to be reliable.

The "one-day maximum" and "end of assignment" data appear more promising. These analyses found fewer than two outliers per week. That rate is similar to the rate for the content-based focused reinterview. There is minimal overlap between the outliers from these two analyses. Therefore, the outliers from both analyses would be good candidates to send out for the focused reinterview.

The interviewers identified in these analyses coincide very little with those identified in the focused reinterview analysis. None of the "end of assignment" outliers were identified by the focused reinterview, and only about 21.0 percent of the "one-day maximum" outliers overlapped with the focused reinterview. If the reinterview check shows the non-overlapping interviewers have falsified data, these variables will add to the effectiveness of the focused reinterview.

References

- Biemer and Stokes (1989), "Optimal Design of Quality Control Samples to Detect Interviewer Cheating," *Journal of Official Statistics*, Vol. 5:1, pp. 23-39.
- Grant, E.L. and Leavenworth, R.S. (1988), <u>Statistical</u> <u>Quality Control</u>, sixth edition, McGraw-Hill, New York, NY.
- Hood, C.C. and Bushery, J.M. (1997), "Getting More Bang from the Reinterview Buck: Identifying 'At Risk' Interviewers," *Proceedings from Section on Survey Research Methods*, American Statistical Association, 1997, pp. 820-824.
- Schreiner, I., Pennie, K., Newbrough, J., "Interviewer Falsification in Census Bureau Surveys," *Proceedings from Section on Survey Research Methods*, American Statistical Association, 1988, pp. 491 - 496.
- Wetzel, A.J. (1993), "Falsification by Field Representatives 1982 - 1992," Census Bureau internal Memorandum.

Table 1. Outlier Assignments Identified by Date Stamp Analysis

	Number Flagged	Percent Flagged	Also in Focused Reinterview	Average Number of Outliers per Interview Period
One-Day Maximum	29	1.2%	6	1.7
Top Three Days	135	5.7%	15	7.9
End of Period				
Ineligible Cases	2	0.1%	0	0.1
Interviews	9	0.4%	0	0.5

Notes: • The data in Table 1 are based on 2,387 assignments checked.

• Beginning with Quarter 2 Week 6 of the 1999 NHIS, outlier FRs identified by these analyses are being sent out for the focused reinterview check.