GETTING MORE BANG FROM THE REINTERVIEW BUCK: IDENTIFYING "AT RISK" INTERVIEWERS

Catherine C. Hood, John M. Bushery, Bureau of the Census
Catherine C. Hood, Bureau of the Census, Demographic Statistical Methods Division, Washington, DC 20233

Key Words: Outliers, Reinterview, Statistical process control

We are conducting a special "focused reinterview" study for the Census Bureau's household surveys. We want to find a more effective way to identify interviewers who falsify data. Interviewers falsify data when they knowingly deviate from specified interviewing procedures to avoid interviewing or properly classifying units.

The goal of this research is to develop methods and tools, such as statistical process control (SPC), for the Census Bureau's regional offices to focus the quality control reinterview on outlier interviewers. We look for interviewers whose work contains unusual distributions, for example, too many units without a telephone number. We not only look for interviewers who stand out for a particular variable, but for interviewers who have a pattern of unusual answers for several different variables.

One difficulty in finding outlier interviewers is variability in the population. Covariates from the 1990 Census help control for variation in the population and highlight variation between interviewers. They can improve efficiency by limiting false alarms and reducing the number of focused reinterviews.

Early results of this study in the National Health Interview Survey (NHIS) suggest that the focused reinterview can be an effective component of an interviewer quality control program. By targeting outlier interviewers, the focused reinterview seems to detect more cases of falsification per interviewer checked. Each year the random reinterview checks about 400 interviewers and detects an average of 1.9 confirmed falsifiers. So far the focused reinterview has checked 83 interviewers and detected three confirmed falsifiers. Three more suspected cases are still under investigation.

Falsification Model

Why would an interviewer falsify data? More than fifty years ago, Crespi (1945) identified several reasons interviewers might falsify. Among the reasons he gave are

- long questionnaires
- complex or difficult questions
- personal or intrusive questions
- unreasonable deadlines and large workloads
- bad roads, bad weather, and bad neighborhoods

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.

The Census Bureau maintains high standards for response rates -- a major component of interviewer performance evaluations. This emphasis on high response rates may tempt some interviewers to falsify data if too many of their sample cases refuse to participate or can't be reached.

We hypothesize a general falsification model --

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

In practice, this model means interviewers who falsify will tend to:

- List eligible units as ineligible. Examples of ineligible units: vacant, demolished, used for business or storage, occupied entirely by people in the Armed Forces, or occupied entirely by people with usual residence elsewhere.
- Screen out cases as not having the characteristic of interest. For example, in the NHIS, some households occupied by Whites are screened out of the sample to achieve an oversample of Blacks and Hispanics.
- Take the shortest path through the questionnaire. They would tend to answer "No" to questions that otherwise would lead to a battery of more detailed questions.
- Omit the telephone number. This makes reinterview more difficult. For recurring surveys, it can also keep the case from being sent to the Telephone Centers in subsequent months.

The Census Bureau's Random Reinterview Program

Fortunately, falsification at the Census Bureau is very rare. The Bureau finds only one in 200 interviewers falsifying data in its ongoing surveys. The rate is higher for one-time surveys which employ many new-hire interviewers.

The reinterview program is one reason the falsification rate is so low. The Census Bureau conducts random reinterviews for all of its household surveys. Because the random reinterviews are unpredictable, the program is a good deterrent against falsification. We conduct quality control reinterview only on cases interviewed in the field. (Because the Bureau's Computer Assisted Telephone Interview Centers monitor interviews, they don't need a reinterview to detect and deter falsification.) Random reinterviews discover approximately three-fourths of the

falsification detected. However, since falsification is so rare, the random reinterview is expensive for the number of cases detected.

In 1982 the Census Bureau began a long-term study to learn more about the characteristics of falsification incidents. This study found that fabricating an interview is the leading type of falsification. Listing an occupied unit as vacant was also very common (Biemer and Stokes (1989) and Wetzel (1993)).

The study also revealed a relationship between the interviewer's tenure and both the proportion of the assignment falsified and the type of falsification detected. New interviewers, with less than five years of experience, have a higher chance of being detected falsifying data. When they falsify it is usually a relatively high proportion of their assignments and they tend to fabricate entire interviews. Interviewers with five or more years of experience usually falsify a smaller proportion of their assignments and tend to classify eligible units as ineligible.

This study led to a more efficient sample design for the random reinterview program. Since new interviewers tend to falsify more often than experienced interviewers, they are checked more frequently. However, since the new interviewers tend to falsify a large proportion of their assignments, checking a smaller number of households can still provide a reasonable probability of finding falsification. The Bureau checks experienced interviewers less frequently than new interviewers. But because they tend to falsify smaller portions of their assignments, checking more households in the assignment increases the probability of detection.

Automating the "Sniff Test" -- Focused Reinterview

With paper and pencil surveys, field clerks and supervisors in the Census Bureau's 12 regional offices occasionally detected suspicious work while processing questionnaires. Field staff often refer to this as the "sniff test." Repetitive answers on several questionnaires and overly meticulous questionnaires are two indicators of possible falsification found during clerical processing in the regional offices. With Computer-Assisted Personal Interviewing (CAPI), survey supervisors have much less information to gauge interviewers' performance. Without paper forms, field supervisors need a way to monitor the data from the computerized questionnaires. Focused reinterview for CAPI interviews can automate the "sniff test."

The Census Bureau took a step toward focused reinterview in 1993 with the National Survey of College Graduates (NSCG). Respondents for this survey were selected from people who reported having a bachelors' degree in the 1990 Census. Response error and imputation in the Census caused about 3 percent of the reported bachelors' degrees to be false. NSCG field interviewers classified "no bachelors' degree" respondents as out-of-

scope and terminated the interview. Misclassifying eligible respondents as "no bachelors' degree" was an easy and obvious way to falsify an interview. In addition to the random reinterview, Census developed a procedure to select interviewers for a special reinterview of "no bachelors' degree" cases. The two falsifiers found were detected because they returned an unreasonably large number of "no bachelors' degree" questionnaires.

Census used another characteristic to focus reinterviews in the 1995 American Travel Survey (ATS) and the 1996 Survey of Fishing, Hunting, and Wildlife Associated Recreations (FHWAR). Since most random reinterviews are conducted by telephone to save money, interviewers who falsify data would be less likely to provide the respondent's telephone number. Census selected interviewers with a large proportion of cases with no telephone numbers. The regional offices checked six interviewers in the ATS and nine interviewers in the FHWAR. One of the six ATS interviewers had falsified interviews. None of the FHWAR interviewers were found to have falsified. However, one of the nine was terminated for not using the laptop computer. These early successes encouraged our current research.

The NHIS is a good candidate for falsification research. It uses a long questionnaire with difficult and personal questions. The NHIS asks a screening question that eliminates some White households. Several questions lead to very long and complicated paths in the questionnaire. The shortest path through the interview, except ineligible units, is a one-person family. This person would be a white nonsmoker, no health problems and no health insurance.

As indicators for possible falsification, we looked at unusually high proportions for the following four variables:

- Ineligible Unit Rate
- Nonminority Screening Rate
- Short Interview Rate
- No Telephone Number Rate

Technical Issues

Separating Interviewer Characteristics from Population Variability

One difficulty in finding outliers is variability in the population. We used covariates whenever possible to control for variation in the population and highlight the variation between interviewers. For example, variability within the population made it difficult to determine when an interviewer screened out too many cases as nonminority. The NHIS sample stratification provided a convenient covariate to control for population variation -- the expected nonminority rate from the 1990 Census.

In April we began using covariates from the 1990 Census long form data. We used these proportions:

- occupied and vacant units, both seasonal and nonseasonal
- persons in the Armed Services (ineligible for the NHIS)
- one-person and no-children households (for the short interview rate)
- units with no telephone

We sum data to the smallest geographic area possible. This allows us to use Census data for a county and sometimes segments within a county. We then can compare the rates from the interviewer's assignment with the rates from the Census data within the area.

Consider the nonminority-screen-out rate to illustrate how we used covariates. Rather than monitoring an interviewer's actual screen-out rate, we monitor Δ, the difference between the rate and the rate we expect for the interviewer's assignment.

Let $\Delta_i = p_{ai} - p_{ei}$, where

p_{ai} = actual rate screened out by interviewer i (in a given regional office)

= number of households screened out ÷ number of households in assignment

p_{ei} = expected rate to be screened out by interviewer i

= expected number of households screened out ÷ number of households in assignment

Expected number of households screened out

$$= \sum_{strata} \text{(stratum nonminority rate)} \times \text{(number of households in assignment in stratum)}$$

Control Limits

We use statistical methods similar to SPC to focus the reinterview on outlier interviewers. As in SPC we calculate control limits based on the natural variability of the process. Any observation exceeding these limits signals a possible problem with the process. We consider any interviewers outside the control limits as outliers and flag their cases for focused reinterview.

Before we started using covariates, we relied on the usual P Charts. When we began to use covariates, our analysis became more complicated. Now we want to monitor Δ instead of p_a. The household-level Δ's (the household characteristic minus p_e) have the same variance theoretically as a binomial variable, but they have neither a binomial nor a normal distribution. Therefore,

conventional P Charts are no longer appropriate. Nor are X-bar Charts.

We finally settled on computing the upper control limit as

$$\text{UCL}_\Delta = \min\left(\bar{\Delta} + 3\sqrt{\frac{s}{n_i}}, \ 1\right), \text{ where}$$

$\bar{\Delta}$ is the average difference between the actual and expected rates for the regional office,

s is the sample variance among the Δ_i for the regional office, and

n_i is the sample size for interviewer i

This method of computing the variance ignores the "logical groups" of the interviewer assignments that would be used in the usual X-bar Charts. This formula tends to produce fewer out-of-control signals.

Multivariate Control Limits

We also want to monitor simultaneously all the variables. Multivariate control charts are based on the Mahalanobis distance -- the standardized distance of each multivariate case to the vector of the process means of all the variables. With correlated variables, a process can be in control for each variable separately, but from a multivariate standpoint, the process can be out of control when the correlation between the variables is taken into account. For more information on the Mahalanobis distance, see Alt (1988), Bienias (1995), and Ryan (1989).

The Mahalanobis distance is essentially one multidimensional region. A serious problem with the Mahalanobis distance in our application is that it does not have upper and lower control limits. We're not interested in whether interviewers screen out too few households as nonminority, just whether they screen out too many. We can't make that distinction using the Mahalanobis distance. We have to examine all the pair-wise scatter plots to decide whether an out-of-control Mahalanobis distance signals the need for a focused reinterview. Cleveland (1993) provides more information on using graphs to find multivariate outliers.

Unfortunately, examining plots is too labor-intensive. The regional offices do not have the resources to do it on a production basis. We want a system that easily would tell the regional offices whether someone should be in reinterview or not. We decided to compute standardized scores for each variable. We add the scores for all four variables. If this combined score is greater than 1.5 (an arbitrary upper control limit), then we flag the interviewer as a multivariate outlier. This method gives us a convenient way to determine whether to reinterview an interviewer, even though it doesn't take the correlation between variables into account.

Early Results for the NHIS Focused Reinterview

Focused reinterview for the NHIS began in February 1997. We have 14 weeks of reinterview forms back from the regional offices. We used SPC methods without covariates to select interviewers during the first eight weeks and included covariates for the next six and all subsequent weeks.

Early in the study, many cases were flagged for "ineligible units" and "no telephone." Fourteen of these cases proved to be in resort areas where many units were in fact ineligible and had no telephone. When we re-ran the analysis, this time using covariates, none of these interviewers were flagged. This result confirmed our hypothesis that covariates would reduce the number of false alarms.

So far, we have focused reinterview results for 83 interviewers. We found seven interviewers (8.4%) with possible falsification: four in the first eight weeks and three in the next six weeks. Table 1 suggests that covariates might reduce false alarms, but the sample sizes are too small for statistical significance.

After the reinterview finds indications of falsification, field supervisors investigate further. They check more cases and probe to confirm whether falsification actually occurred.

Although results are still preliminary, the focused reinterview appears to detect falsification at a higher rate than the random reinterview. After checking 83 interviewers, the focused reinterview has detected three cases of confirmed falsification, 3.6 percent. From 1983 through 1996, the random reinterview checked more than 400 interviewers each year and detected 15 cases of confirmed falsification, about 0.2 percent. (Table 2)

The regional offices also found 11 interviewers with procedural violations. They found minor problems, such as entering the wrong outcome code for ineligible households, and more serious ones, such as not using the laptop computer for nonminority screening cases.

Limitations

The focused reinterview has two important limitations. Both limitations make it more difficult to detect experienced interviewers who falsify data selectively.

- It is unlikely to detect interviewers who falsify part of an interview. But neither can the current random reinterview program. The CAPI reinterviews were designed to be a quick-check, quality control reinterview. To minimize respondent burden and costs, each reinterview takes less than five minutes. The reinterview checks only whether an interviewer visited the household, verifies the roster, and checks that the interviewer asked questions about the survey topic.
- It is unlikely to detect interviewers who falsify only a small proportion of their assignments. In order to be checked in the focused reinterview, an interviewer must be a statistical outlier.

Table 1
Number of False Alarms and Falsification Cases for the NHIS Focused Reinterview

	Covariates Used?	Interviewers Checked	False Alarms	Suspected Falsification
Total		83	76 (91.6%)	7 (8.4%)
Weeks 1-8	No	66	62 (93.9%)	4 (6.1%)
Weeks 9-14	Yes	17	14 (82.4%)	3 (17.6%)

Table 2
Focused versus Historical Random Reinterviews for the NHIS

	Time Span	Interviewers Checked	Suspected Falsification	Cleared	Confirmed	Investigation Pending
Focused	14 Weeks	83	7	1	3 (3.6%)	3
Random	Yearly Avg. 1983-1996	400 +	1.9	0.8	1.1 (0.2%)	0

Because falsification is so rare, it may take a few years before we have statistically significant comparisons between the focused and random reinterviews.

Next Steps

Performance-Related Variables

We plan to add variables representing interviewer performance. For example:

- the number of cases transmitted in a day
- the number of times a case was accessed
- the total time spent in the case

Date and time stamps recording each time interviewers access cases and for each section of the interview could help detect partial fabrication. Since the NHIS interview is in several sections it would be fairly easy to complete one section of the interview and fabricate other sections. This type of falsification is difficult to check in any reinterview. If the interviewer has been to the house at all, it is difficult to detect and confirm that only a partial interview was completed.

Current Population Survey (CPS)

The CPS focused reinterview began with one regional office in early August using July survey data. Once the system is running smoothly, we will compute control limits for all Regional Offices every month.

The CPS interviews fall into three categories: CAPI personal visit interviews, CAPI telephone interviews, conducted in the field, and CATI interviews (which require no quality control reinterview). We interview each household eight times. Personal visits are scheduled for interview months 1 and 5. Other months are completed mainly by telephone.

Our initial CPS research will use Month 1 cases only. We will use the no-telephone-number rate and the ineligible-unit rate.

We also plan to investigate CPS supplement noninterview rates as an additional variable. The CPS interview is short but often has supplements to the basic interview. Unlike the basic interview, interviewers can skip the supplement and not be penalized on their performance evaluations. So an interviewer can make the interview shorter by skipping the supplement. We hope to expand the focused reinterview program to look at interview months 2 through 8. The CPS instrument makes it very easy for the interviewer to answer some questions the same as the previous month. We may want to look for unusual consistency between months for the same household.

Finally, we look forward to expanding the focused reinterview program to more surveys.

References

Alt, F.B. and N.D. Smith (1988), "Multivariate Process Control," in P.R. Krishnaiah and C.R. Rao, eds., *Handbook of Statistics, Volume 7; Quality Control and Reliability*, pp. 333-351. Elsevier Science Publishing Company: New York, NY.

Biemer and Stokes (1989), "Optimal Design of Quality Control Samples to Detect Interviewer Cheating," *Journal of Statistics*, Vol. 5:1, pp. 23-39.

Bienias, J.L. (1995). "Multivariate Outlier Detection Methods for the Quarterly Financial Report," *Proceedings of the 1995 Bureau of the Census Annual Research Conference and Seasonal Adjustment Workshop*.

Cleveland, W.S. (1993). *Visualizing Data*. Hobart Press: Summit, NJ.

Crespi, L.P. (1945). "The Cheater Problem in Polling," *Public Opinion Quarterly*, Winter, pp. 431-445.

Ryan, T.P. (1989). *Statistical Methods for Quality Improvement*, pp. 216-223. Wiley: New York, NY.

Wetzel, A.J. (1993). "Falsification by Field Representatives 1982-1992," Census Bureau Memorandum for Paula Schneider.