AUTOMATING THE FOCUSED REINTERVIEW

John S. Hilton, John M. Bushery, U.S. Bureau of the Census
John S. Hilton, Census Bureau, Room 2228/SFC-2, Washington, D.C. 20233

Key Words: Automation, Reinterview, Outliers

INTRODUCTION

Brief History of Reinterview at the Census Bureau

The Census Bureau uses a reinterview program in its ongoing surveys to detect and deter falsification by interviewers. This program has two components: a random reinterview and a supplemental reinterview. The random reinterview usually consists of randomly selecting interviewers and some of their cases for reinterview, while program supervisors in the field offices purposively select cases to reinterview in the supplemental system.

In 1997 the Census Bureau added an experimental "focused reinterview" to the Bureau's reinterview program for the National Health Interview Survey (NHIS). The focused reinterview (FRI) attempts to more effectively identify interviewers who falsify data by using methods and tools, such as statistical process control (SPC), that focus the reinterview on outlier interviewers (Hood, 1997). It is a partial answer to the concern of the field offices no longer being able to see the interviewer's work, and it capitalizes on the more timely transmission of Computer Assisted Personal Interview (CAPI) data.

All components of the reinterview program are administered by telephone (when possible) to control costs and seek to confirm a few key items obtained in the original interview. Historically the Census Bureau has found about one in 200 interviewers falsifying data in its ongoing surveys (Wetzel, 1993). The random reinterview program detects about 75 percent of this falsification. The remaining 25 percent is detected in a variety of ways, usually resulting in using the supplemental reinterview system to confirm falsification. In its first year, the focused reinterview checked about 200 interviewers and detected three interviewers falsifying data.

Focused Reinterview

The focused reinterview uses control charts to identify those interviewers who stand out for a particular variable and those who have a pattern of unusual answers for several different variables. These variables are best selected by those familiar with the intricacies of the surveys; the more knowledge one has of a survey the better he or she is at selecting variables that indicate possible falsification. The NHIS, Current Population Survey (CPS), and Survey of Program Dynamics (SPD) are the three surveys that currently use a focused reinterview. Table 1 shows the variables we monitor that might indicate falsification for these three surveys.

Table 1. FRI variables for NHIS, CPS, and SPD

Variable	Survey		
	NHIS	CPS	SPD
No phone number rate	X	X	X
Non-interview rate	X	X	
Rate of screened out cases	X		
Rate of short interviews	X		

Focused reinterview works better in some surveys than others, and while used in three different surveys, it is still experimental in nature. Falsification is rare, so it may be several years before we have enough data to make statistically significant comparisons between the random and focused reinterviews.

Because most of our experience in the focused reinterview comes from the NHIS, we will describe the process for that survey. We receive original interview data files that we analyze to identify interviewers who are outliers. We then print reinterview forms for these interviewers' cases and mail them to the field offices. The field offices assign reinterviewers to conduct the focused reinterview and mail or fax the forms back to us upon completion. As we can see, the focused reinterview is a paper and pencil interview (PAPI) even though the original NHIS uses a CAPI system.

Currently the FRI is still experimental so very little of the process has been automated. Many of the tasks require a person to be physically present, thus illness and vacations can cause delays in the focused reinterview. To move to a production system the process must be automated as much as possible. Automation will make the focused reinterview a more effective component of the reinterview program.

Statistical Techniques

Methods similar to SPC techniques are used in the focused reinterview to determine if an interviewer is outside the control limits for the variables of interest. To try to control for variation, each interviewer is compared only to other interviewers within the same region. Also, we use covariates whenever possible to further control for variation in the population. The net effect is interviewers are considered outliers only if their work is significantly different from what we would expect. Thus an interviewer with an assignment in a resort area would not be considered an outlier simply because many of the cases did not have a telephone number (Hood, 1997).

Supplemental Reinterview System

The Census Bureau will soon automate the existing supplemental reinterview system so that the field offices can reinterview cases in the current assignment period. Currently the field offices have to make a special request of headquarters when they want specific cases reinterviewed; it is apparent how this system can be cumbersome and inefficient at times. The new supplemental reinterview system, which began in June, 1998 in the SPD, will complement the random and focused reinterviews and give the field offices an additional tool for producing high quality interviews.

ADVANTAGES OF AUTOMATING THE FOCUSED REINTERVIEW

Cost Effective

Some cost savings may potentially result in a reinterview program over time as automation occurs. We expect most of these savings to be in headquarters rather than in the field offices. NHIS is a weekly survey so about once a week the focused reinterview is processed. From start to finish it takes about one day for headquarters personnel to complete their part of the weekly focused reinterview; this length of time will diminish as the process becomes more automated. Human intervention will also diminish as the process automates, resulting in the expected cost savings.

Improved Timeliness

Some of the reasons timeliness is important in a survey environment are: field supervisors get more timely feedback on their interviewers, less time has elapsed for respondents who are reinterviewed so they should have a better memory of the original interview, and the reinterview process in general is more efficient simply due to the quicker turnaround. As we have begun the automation process we have realized a savings of about 10 days. The focused reinterview uses original interview data to determine outliers, and changing how we obtained these data resulted in the 10 day savings. Further automation should save another seven days or so, mostly by eliminating the mailing of FRI forms to the field offices. Rather than mail paper forms to the field offices, we will electronically notify them of cases to select for FRI and when completed they will transmit the cases back to headquarters.

STEPS IN THE AUTOMATION PROCESS

We have identified three steps in automating the focused reinterview. Step I involves obtaining data on a daily flow basis; Step II is converting the FRI from PAPI to Computer Assisted Interviews (CAI) using the new supplemental system; and Step III is producing reports showing which interviewers are flagged as outliers. We discuss each step in greater detail below.

Step I -- Automated Data Processing

The main task in Step I was obtaining original interview data on a flow basis. This change meant a new data supplier as well as different datasets. While this was a conceptually simple task, changing programs to accommodate the new data was not trivial. When the focused reinterview first began in 1997, we would get a final dataset of original interviews. At times there was a delay in getting this dataset because of something called "closeout." Before a dataset was considered complete, all 12 of the field offices had to closeout which essentially signified that all the cases for that sample week were accounted for. So if only one field office had a few remaining cases, the complete dataset could not be compiled and the reinterviews were delayed.

A clear advantage to receiving data on a flow basis is that we don't necessarily need 100 percent of the data to ensure reliable results and thus don't have to wait for all 12 field offices to closeout. For three weeks in December 1997 we had data from our old and new sources, allowing us to do the FRI on both sets of data. We knew how much of the partial data was needed by noting how complete it was when it yielded the same results as the complete data. The percent of partial data needed to obtain the same results as the complete data were 99.6 (week 1), 89.6 (week 2), and 83.1 (week 3). Because this is only three weeks of data and we wanted to be conservative, we decided to conduct the FRI analysis when we had 95 percent of the original interview data.

Knowing when we had obtained 95 percent of the data turned out to be more difficult than anticipated. To calculate the percentage, we needed to know how many cases were sent out each week. The seemingly simple task of obtaining the number of cases took significantly more time than we expected. From this we learned that careful attention to small details and anticipating problems can save much time and effort at a later date.

Another challenge we faced was in simply accessing the data. One of the first steps in analyzing any dataset is being able to read it and manipulate it. The NHIS data contained several records for each household or case; reading this type of data presents some difficulty but can be done in a straightforward manner. NHIS is a complex survey with many details that can be difficult to communicate to all the right people. When changes are made to the instrument or the data structure or any other aspect of the survey, effective communication between the different groups working on the survey is absolutely essential. Otherwise it can be rather difficult to use the data.

One aspect of focused reinterview different from regular reinterview is that we cannot control how often an interviewer is flagged for focused reinterview. We do not want to burden the field staff by asking them to reinterview an interviewer who they have recently reinterviewed (either in focused or regular), so it is important to keep track of which interviewers we have sent to the field for focused reinterview and when we sent them. Determining how long one must wait before reinterviewing an interviewer who has again fallen into focused reinterview is important. We decided to keep track of the previous two quarters of focused and random reinterview activity. So if an interviewer is flagged for focused reinterview in the current period but was reinterviewed in the past two quarters, they would not be sent to the field office to be reinterviewed.

Step II -- Automated Instrument

NHIS and CPS use paper reinterview forms to conduct the focused reinterview. For NHIS we print the forms at headquarters and mail them to the field offices; the CPS reinterview forms are stored electronically on a secure server and the field offices can then access and print the forms themselves. Step II involves replacing the FRI paper forms with a CAI instrument.

A major advantage of using a CAI instrument is the time saved over using the paper forms. Rather than mailing forms to the field offices or having the field offices print the forms, we can transmit the FRI cases directly to the reinterviewer. Once the reinterviewer completes the FRI he or she can transmit the cases back to headquarters and bypass the field office entirely. This should result in a savings of about one week. It should also result in more accurate reinterviews as the instrument uses internal consistency checks and edits to help ensure accuracy.

The new supplemental reinterview system will allow us to conduct the focused reinterview with CAI technology. One of the basic changes made by the new system is the field office can select virtually any current case and reinterview it. Previously the field offices had to request that an interviewer's cases be made available for supplemental reinterview, but it was only the next assignment period set of cases that were eligible for supplemental reinterview. The new supplemental system allows the field office to select cases as soon as falsification is suspected, allowing the field office to check the actual cases for which falsification is suspected. The focused reinterview selects certain cases for the outlier interviewers, and a list of these cases will be sent to the field offices for them to reinterview. The 1998 SPD focused reinterview will be the first survey to test the performance of this part of the automation process.

Step III -- Automated Analysis

The first part of Step III is turning all of the SAS programs we currently use to do the focused reinterview over to our programming branch. They will take the various programs and make them into a seamless, automated system. The end product of this system will be a report of the interviewers flagged as outliers and the cases to be included in the reinterview.

Once the outlier report has been produced, we will review it to make sure there are no obvious reasons why we should not include an interviewer in the focused reinterview. The programs automatically ensure that no interviewers are sent out who have been reinterviewed in the last two quarters. They also make sure there are enough cases for each flagged interviewer; enough in this case is five. Our review does not look for anything specific, it mostly checks for obvious reasons that would indicate the cases should not be sent out. Once we review and approve the list of flagged interviewers, we transmit the cases to the reinterviewers.

LIMITATIONS AND POTENTIAL PROBLEMS

Human Intervention Still Necessary

At the completion of Step III, the focused reinterview system will be as automated as possible, but not

completely automated. Some human intervention will still be needed to monitor the system and review the reports before sending them to the field offices. Occasionally the computer programs should be reviewed, and modifications will be necessary as requirements change. For example, the control limits may need to be adjusted or the variables of interest may change.

Before sending a list of the cases to reinterview to the field offices, the list must be reviewed for cases that should not be included. Some areas where the cases are located are obvious resort or dormitory areas, so a lot of vacant units or units with no telephone numbers are expected. The analyst noticed one set of cases located in the Poconos area of Pennsylvania, a popular winter resort area. The analyst recognized another set as dorms at a university, as she was an alumnus of that university. Each time forms were not sent out for reasons such as these, the field offices saved valuable resources by avoiding ineffective reinterviewing.

High Development Costs

Developing a focused reinterview system can potentially incur some high costs. Research must be conducted on which variables might indicate falsification, what covariates to use, and the statistical methods to employ. Programs must be written that incorporate the research into a viable reinterview system. Automating the system will also take time and expend resources that add to the cost of the entire system. One thing to keep in mind is the cost of producing a new, automated system may be higher than administering the new system, especially for one time surveys. Thus costs may initially be higher in headquarters than in the field offices.

Communication

An obvious potential problem that we cannot stress enough is the need for effective communication between all the interested parties. The reinterview system for NHIS involves at least five separate teams. These teams reside in different divisions and have different management and communication structures. But for the reinterview system to work properly and effectively, the lines of communication between the teams and the responsibilities of each team must be very clearly defined.

General Limitations

With or without automation the focused reinterview has two important limitations. Both limitations make it more difficult to detect experienced interviewers who falsify data selectively.

- The FRI is unlikely to detect interviewers who falsify part of an interview. The CAI reinterviews were designed to be a quick, quality control reinterview. 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. To be checked in the focused reinterview, an interviewer must be a statistical outlier.

These limitations point out the need to have both a focused reinterview and a random reinterview. These two types of reinterview complement one another rather than replace one another. An effective reinterview system will use both types of reinterview.

References

Hood, C.C. (1997). "Getting More Bang from the Reinterview Buck: Identifying 'At Risk' Interviewers," *Proceedings of the 1997 Joint Statistical Meetings*.

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