Evaluation of the Quality Assurance Falsification Interview used in the Census 2000 Dress Rehearsal

Elizabeth A. Krejsa, Mary C. Davis, Joan M. Hill, U.S. Census Bureau
Elizabeth A. Krejsa, U.S. Census Bureau, PRED_9200, Washington, D.C. 20233-9200

1. ISSUES

In order to reduce errors in surveys, quality assurance (QA) programs are often instituted. The purpose of the falsification interview is to evaluate interviewer's work and identify those who are either falsifying data or are not following correct procedures. One way interviewers are targeted for the falsification interview is by examining the original interview characteristics (such as the length of the interview), respondent characteristics (such as age), or interviewer characteristics (such as being a new interviewer). Tolerance levels are used to determine if an interviewer exceeds acceptable limits.

There are three important issues relating to falsification interviews. First, most interviewers do not falsify data, thus finding the interviewer who does is like searching for the "needle in the haystack." Second, because resources are being used to conduct the falsification interviews, the interview must be effective compared to other QA programs. There is also additional respondent burden.

To address these issues the following questions are answered:

1. Does targeting for falsification interviews identify significantly more falsified cases than a random sample?
2. Do the QA supervisors understand the targeting reports and use them correctly? Do the tolerance levels need to be adjusted to meet workload projections?
3. Is the QA process successful in detecting falsified cases?

2. METHODOLOGY

The Integrated Coverage Measurement/Post Enumeration Survey (ICM) is a personal visit interview (PI) conducted using a Computer Assisted Personal Interview (CAPI) instrument. The survey was conducted as part of the Census 2000 Dress Rehearsal in Sacramento, CA, Menominee County, WI and Columbia, SC and the eleven surrounding counties. To help ensure that the data collected in the PI were valid, a QA falsification interview was conducted when a case was:

1. selected as part of a five percent systematic sample or
2. targeted based on predetermined criteria.

A QA interviewer returns to the household and determines if the respondent, or someone else in the household, was interviewed. If the respondent was not interviewed during PI, then the QA interviewer conducts the PI interview with a CAPI instrument. If the respondent or someone else in the household was interviewed, then it is assumed that the case was not falsified and the PI interview is not conducted again.

Three types of targeting reports were developed to assist the QA staff in targeting cases. The reports contain statistics believed to be good indicators of problem interviews and are intended to assist QA supervisors in targeting suspicious cases for QA falsification interviews. The systematically selected cases are not eligible to be targeted. The three targeting reports are explained below.

2.1 Field Representative Outlier Report

The Field Representative Outlier Report identifies field representatives (FRs) who are outliers for a number of variables possibly related to falsification and data quality. Tolerance levels are predetermined for some variables; for others, the tolerance level depends on the average for the geographic area. The outlier variables with predetermined tolerance levels are:

- number of interviews that are less than two minutes
- number of cases completed between 10 p.m. and 8 a.m.
- days with more than thirteen interviews

The outlier variables with tolerance levels based on the average for the geographic area are:

- percent of cases with no phone number
- percent of cases with missing outmover data (where a noninterview occurred for Census Day but an interview was completed or a vacant household was found on interview day for the PI)
- percent vacant (percent of vacant housing units and housing units that do not exist or are no longer used

for residential purposes, or is now a business, on interview day)

- percent proxy (cases completed by a respondent who is not a household member)
- percent partial interviews (cases in which some data are missing for at least one person in the household)

QA supervisors use the Field Representative Outlier Report to identify FRs who may be falsifying data or may not be following correct procedures. Three variables (percent proxy, percent partial, and missing outmover) were used as quality indicators to help identify FRs who may not be following correct procedures. All other variables were used as falsification indicators. One set of reports is produced for cases completed by telephone and one set of reports is produced for cases completed in person.

Tolerance levels are calculated for a specific geographic region to determine outliers. It is assumed that cases within a geographic region are of similar difficulty for the FR and have similar characteristics.

Moderate Outlier calculations were based on a one-sided 95% confidence interval:

$$p_{FR} > p_{pop} + 1.645*(\sigma/sqrt(n_{FR})) \quad AND$$
$$p_{FR} < p_{pop} + 2.333*(\sigma/sqrt(n_{FR}))$$

Severe Outlier calculations were based on a one-sided 99% confidence interval:

$$p_{FR} > p_{pop} + 2.333*(\sigma/sqrt(n_{FR}))$$

where:

- FR is the Field Representative
- p_{FR} is the FR proportion of cases
- p_{pop} is the proportion of cases currently completed and sent back from field within a geographic area (population)
- n_{FR} is the number of Field Representatives in the given population
- $\sigma=sqrt(p_{pop}(1-p_{pop}))$ is the population standard deviation

The purpose of the two outlier calculations was to allow supervisors to focus on only the severe outliers if the workload for the QA falsification interview became too large when both severe and moderate outliers were targeted.

2.2 Respondent Name Report

The Respondent Name Report enables the QA supervisor to browse the entries for respondent names made by each FR. The aim is to look for indications of possible falsification, such as names of famous characters/people or multiple respondents with the same name, for example, Snow White or Donald Duck.

2.3 Not Enough QA Cases Report

The Not Enough QA Cases Report enables the QA supervisor to identify FRs who have completed and sent back at least ten interviews but have no cases selected for the QA falsification interview. This report is especially helpful at the beginning of the QA process when there are not a lot of cases to generate the outlier reports.

2.4 Identifying Falsification

Falsification is identified through a series of phases. In the first phase a flag is set in the QA CAPI instrument to indicate if the respondent or someone else in the household had not been interviewed. In the second phase, after the QA falsification interview, the FR indicates whether he/she felt that the original interview was falsified based on further comments by the respondent. In the third phase, these two flags are combined to produce a suspected falsification flag as follows:

If the QA CAPI falsification flag (phase one) is...	If the FR falsification flag (phase two) is...	Then the suspected falsification flag is...
Falsified	Falsified	Falsified
Falsified	Can't Determine or Blank	Falsified
Can't Determine	Falsified	Falsified

All other combinations are set as either 'Not Falsified' or 'Can't Determine'.

If in the third phase an FR is suspected of falsification, the supervisors may send additional cases from that FR for the QA falsification interview to possibly determine a pattern of falsification.

In the fourth phase, supervisors review cases returning from field with a suspected falsification flag of falsified and determine whether he or she felt the case had been actually falsified. They base their decision on notes from the FR, the reliability of the respondent, and by reviewing other cases from the same FR that have been in the QA

falsification interview process. The results of the inquiry were used to determine confirmed falsification. Once an FR is confirmed to have falsified cases the FR's workload is reinterviewed.

3. LIMITATIONS

The following limitations affect interpretation of the results of the analysis.

- Since falsification is a rare event, sample sizes obtained in the QA operation may not be sufficient to determine if the differences between the randomly selected sample and the targeted group are statistically significant. Therefore, conclusions and/or recommendations from this evaluation are based on both quantitative statistical measures and anecdotal or qualitative evidence.

- Although the sites (Sacramento, CA; Columbia, South Carolina; and Menominee, WI) were chosen to represent situations found throughout the country, results of this evaluation cannot be generalized to any area beyond the three sites.

4. RESULTS

Based on the falsification flag alone, in Menominee, of the 12 PI FRs, only one FR was suspected of falsifying a case. In Sacramento, of the 218 PI FRs, 30 were suspected of having falsified cases with an average of 1.21 falsified cases per FR suspected of falsification. In South Carolina, of the 249 PI FRs, 32 were suspected of having falsified cases with an average of 1.56 falsified cases per FR suspected of falsification. However, only three FRs in one site and none in the others were confirmed to have falsified cases, thus re-enforcing that falsification is a rare event.

4.1   Reason Case Sent to QA and Percent of Falsified Cases

Table 1 indicates the number of cases sent to QA for each method of selection, cases suspected of being falsified based on the falsification flag, and cases confirmed to be falsified through supervisor review. It can be seen that the targeting procedures identified more possibly falsified cases and confirmed falsified cases than did the systematic sample.

Table 1: Cases Sent to QA and Cases Falsified

Menominee

Reason sent to QA	Falsified		
	Cases	Suspected	Confirmed
Targeted	81	1	0
Sampled	32	0	0
Total QA	113	1	0

Sacramento, CA

Reason sent to QA	Falsified		
	Cases	Suspected	Confirmed
Targeted	875	23	0
Sampled	821	16	0
Total QA	1696	39	0

Columbia, SC and Surrounding Counties

Reason sent to QA	Falsified		
	Cases	Suspected	Confirmed
Targeted	781	33	10
Sampled	853	17	1
Total QA	1634	50	11

When examining the targeted cases we found that in all three sites, the "Not enough QA cases" category was used most often to send a case to QA. In Menominee, only one case was suspected as being falsified, therefore, no meaningful comments can be made about the targeting variables.

In Sacramento, the Respondent Name Report and the "Combination" category were most often used, after "Not enough QA cases", to send a case to QA. The "Respondent Name Report" and the "Not Enough QA" category identified the most cases suspected of falsification.

In South Carolina, the "Length of interview" category was also used heavily to target. In South Carolina, the "Length of Interview" category and "Combination" category identified the most cases suspected of falsification.

4.2 Statistical and Practical Considerations

Statistical Considerations. To examine if the targeting procedure identified significantly more suspected falsified cases than the systematic sample, we tested the null hypothesis that the difference between the targeted suspected falsified proportion of total cases and the systematic sample suspected falsified proportion of total cases is less than or equal to zero (for alpha equals .10).

In Menominee, WI the difference between the targeted percent and systematic percent is 1.23% with a standard error of 1.95% and a p-value of .264. Therefore, the targeting procedure did not find significantly more suspected falsified cases than systematic sampling.

In Sacramento, CA the difference between the targeted percent and the systematic percent is .69% with a standard error of .73% and a p-value of .171. Therefore, the targeting procedure did not find significantly more suspected falsified cases than systematic sampling.

In Columbia, SC and the surrounding counties the difference between the targeted percent and the systematically sampled percent is 2.24% with a standard error of .85% and a p-value of .004. Therefore, the targeting procedure found significantly more suspected falsified cases than systematic sampling. The difference is also significant for the confirmed falsified cases, where the targeting procedure led to confirming significantly more falsified cases than systematic sampling.

Practical Consideration. There are additional costs associated with conducting a targeting review versus the systematic sample. The cost to conduct the QA falsification interview (pay for the FR) and the cost for a QA supervisor to handle a falsified case is the same for either method. The additional cost for the targeting review is the QA supervisor's review of the targeting reports. As shown below, this additional cost is relatively small.

In Menominee, WI the annual salary of the QA supervisor reviewing the reports was $36,000. Based on the salary and an estimated fifteen minutes per day spent reviewing the reports, the cost of targeting for the eleven week and three day QA operation was 246.83 dollars.

In Sacramento, CA the annual salary of the QA supervisor reviewing the reports was $39,000. Based on the salary and an estimated two and one-half hours per day spent reviewing the reports, the cost of targeting for the eleven week and three day QA operation was $2,678.30.

In Columbia, SC the annual salary of the QA supervisor reviewing the reports was $61,000. Based on the salary and an estimated one hour per day spent reviewing the reports, the cost of targeting for the eleven week and three day QA operation was $1,675.66.

However, the QA aim is to target five percent in addition to the five percent systematic sample. If targeting was not conducted, the systematic sample would more than double to detect the same level of falsification, thus increasing FR costs. Therefore, it is more beneficial to target based on cost alone.

In all three sites, targeting identified a higher percent of falsified cases than systematic sampling. The difference is significant only in South Carolina, for both the suspected and confirmed falsified cases. Given this and the cost considerations mentioned above, continuing targeting is worthwhile.

4.3 Changing the Targeting Reports

Omitting Variables. The following variables should be removed from the Outlier Report:

* Missing Outmover Data
* Number of days with more than 13 completed interviews

In all three sites, these categories did not assist in targeting cases for QA falsification interview. However, there is no serious impact on the QA operation by leaving categories on the reports that do not target cases.

Modifying Variables. The following describes modifications to the variables that should be made.

* In all sites, the "Combination" category was difficult to interpret for analysis since specific combinations were not listed on the targeting reports. Independent of the reports, it was found that combinations of variables from the targeting reports did not occur often enough to warrant a specific combination category, such as a "Missing Outmover and Partial" category. Thus, the combination category should continue to be used and not modified.

* It appears that "Missing Phone number" is not an accurate predictor of falsification (only 9 cases were targeted and none of them were suspected of falsification), however, when calculating the missing phone numbers per FR, a phone number was classified as missing provided the phone number field contained only blanks. Even though the CAPI instrument made sure that 10 digits must be entered for a phone number, some phone numbers were shown as '1' due to an error occurring after the data collection phase. In the Sacramento, CA site 2.71% of cases had blank phone numbers whereas 16.56%

had a phone number of '1'. In the South Carolina site, 3.26% of cases had blank phone numbers, whereas 11.47% of cases had a phone number of '1'. These cases should have been accounted for in the missing phone number targeting because they are invalid phone numbers. This finding suggests that more cases should have been targeted based on a missing phone number. Thus, the "Missing Phone Number" category should be modified to include invalid phone numbers.

- In all sites, the "Partial Interview" and "Proxy Interview" categories adequately target cases and do not need to be modified. Although used as a quality indicator, the "Partial Interview" category seems to actually be an indicator of falsification and should be treated as such by supervisors.

- In all sites, the data used to calculate the length of the interview for the "Number of complete ICM interviews with length less than two minutes" variable was not accurate. It seems that the PI CAPI software recorded the time when the computer files were opened and closed, rather than recording the actual conversation/interview time. Therefore, the variable in the PI CAPI instrument should be revised.

- In all sites, the "# of Cases complete between 10:00 p.m. and 8:00 a.m." should not be modified. However, the variable used to indicate starting time of the interview was not functioning correctly. Therefore, the variable in the PI CAPI instrument should be revised. If this is not possible, QA supervisors should sample cases in this time range.

- The "Vacant or not a Housing Unit" category was successful in identifying falsified cases.

Adding Variables. In exploring other predictive variables (such as type of basic address and size of household) one possible addition was found. Missing names appear on the Respondent Name Report, however it may be difficult for supervisors to catch slightly large amounts of missing names. Therefore, it is suggested that either a category be added to the Outlier Report or instructions should be given to the supervisors so that they specifically check for missing respondent names while reviewing the Respondent Name Report.

4.4   Office Staff Debriefing

The office staff supervisors felt there was adequate training for the QA falsification interview and that the QA operation was effective. The QA falsification interview identified poor quality of work more often than falsified cases. The threat of QA intimidated FRs and helped deter them from falsifying cases. Based on comments received by the QA supervisors, it appears they understood the reports and used them correctly.

- *Removing the Targeted QA Flag.* Supervisors want to be able to "unflag" a FR from the reports for a specific variable under certain circumstances for a portion of the QA operation. For example, if the FR is working in a seasonal area, they will have high vacancies and thus be flagged. The supervisors would prefer to be able to remove this flag when the indicators do not suggest falsification.

- *Respondent Name Report.* Supervisors said that the Respondent Name report requires a lot of time and that there is too much to look at. The reports were reviewed daily, but not for each FR. Sacramento did one geographic area per day. Because the Supervisors believe the report would be helpful in Census 2000, it should continue to be used.

- *Not Enough QA Cases Report.* Once the FR has had 10% (15% was also suggested) of their cases go through QA they should stop appearing on this screen but continue to appear on the outlier reports. The QA supervisors aim to QA 10% of the FR's cases.

5. CONCLUSIONS/RECOMMENDATIONS

In general, the QA reports were effective in targeting cases for QA falsification interview compared to the systematic sample. The confirmed falsification found through the QA process was low ranging from 0% in Menominee, WI and Sacramento, CA to .06% of the total PI cases in the South Carolina site. QA supervisors seemed to understand the reports and use them correctly. FRs were deterred from falsifying by the threat of the QA process.

5.1   Do the targeting reports identify significantly more falsified cases than random selection?

Based on the results from the test of statistical significance of differences, targeting should continue to be used to identify falsification. Targeting identified significantly more falsified cases in some areas, but did not in others, thus targeting is worth doing at least in some areas.

Based on practical considerations, targeting should continue to be used to identify falsification. There is additional cost associated with reviewing the targeting

reports to identify cases for QA falsification interview. However, had targeting not been conducted, the systematic sample size of 5 percent would more than double thus increasing FR costs beyond the costs of the review of the targeting reports.

5.2 What changes to the QA Model are recommended?

- In addition to edit/validation checks built into the ICM PI CAPI instrument for missing phone numbers, the tolerance level calculations should do further checks for invalid phone numbers. The "Missing Phone Number" category should be recalculated to include phone numbers that were blank as well as any phone number less than ten digits in length. If possible, further checks on phone numbers should also be put in place. For example, a phone number of '2222222222' should be excluded as a valid phone number and be reported as missing. Area codes that begin with a '0' or '1' should also be excluded since no area codes begin with either number.

- Remove the "Missing Outmover" variable from the model if no other operational consideration exists to leave it in.

- Remove the "Days with more than 13 interviews" variables from the model if no other operational consideration exists to leave it in.

- Add a "Missing Respondent Name" variable to the outlier report or instruct supervisors to check for this on the Respondent Name Report. A missing respondent name is defined where a respondent name is blank.

5.3 What changes to QA Procedures are recommended?

- QA supervisors should be sure to carefully monitor QA FRs since there is no quality check on the QA FRs.

- After a QA falsification interview is conducted cases suspected of falsification are returned to the QA supervisor for confirmation of falsification. For accurate analysis, a variable to record the QA supervisors decision in regards to confirmed falsified cases should be created.

6. CENSUS 2000 AND BEYOND

For Census 2000, procedures have been modified so that supervisors are told to look for missing respondent names on the Respondent Name Report. The "Length of interview" variable has been removed from the Targeting Report because of software problems in determining the actual length of the interview. The two variables that were not shown to indicate falsification are being left in the Targeting Reports and the problem with the '1's in the Missing Phone Number has been corrected, however, no further checks for invalid phone numbers are being included.

For other surveys, the use of a targeting interview can be beneficial as shown here. Targeting reduces the amount of random sampling that would need to be conducted and does not inflate costs. Variables specific to the study (such as the Missing Outmover category here) may also be used to further enhance the targeting procedures.

Further study will be conducted on the data received in Census 2000. This may allow us to identify new variables to include in the targeting reports or modify existing variables. Beyond Census 2000 and in other computerized surveys a new technology referred to as Computer Audio Recorded Interviewing (CARI) may be used to identify falsification or interviewing problems. This technology records parts of the interview or the entire interview. This information can then be checked for the presence of more than one voice and confirm that item data recorded in the instrument was actually given to the FR.