

QUALITY ASSESSMENT AND CONTROL STUDIES IN THE AMERICAN TRAVEL SURVEY¹

Lisa Clement, James Hartman, Carol Mylet, Dennis Schwanz, Bureau of the Census
Lisa Clement Room 1661-3, Washington, D.C. 20233

KEY WORDS: Travel, Recall, Response Error, Computer Assisted Telephone Interviewing

1. INTRODUCTION

The Bureau of the Census is conducting the 1995 American Travel Survey (ATS) for the Bureau of Transportation Statistics (BTS) of the U.S. Department of Transportation (DOT). The purpose of the ATS is to determine:

- The number of business or personal trips of 100 miles or more from home taken in calendar year 1995, including the origins and destinations of these trips.
- All transportation services and facilities utilized for the trip, including the transportation used to get to airports, railroad or bus stations, piers or terminals.
- The basic demographic and economic characteristics of the traveler.
- The size of the travel party, the purpose and length of the trip, and modes used for each major intercity corridor.

The ATS consists of a state-based sample design of approximately 80,000 addresses from retired 1980-based Current Population Survey (CPS) sample that were last interviewed in December 1990 to October 1994. The sample size differs by state. This sample was selected to enable BTS to make reliable state estimates.

The survey, which is primarily a telephone interview survey, employed a split-sample design with cases in each CPS primary sampling unit (PSU) being assigned for interview in the Census Bureau's three telephone centers (centralized CATI) or for interview by field representatives (FRs). The cases assigned to FRs could be done by telephone (decentralized CATI) or by personal visit using laptop computers (CAPI). About 45,000 cases were assigned to centralized CATI and 35,000 were assigned to decentralized CATI/CAPI.

In December of 1994, all sampled addresses were sent a letter of notification telling them that they were

selected for the ATS and, following that, a letter of introduction giving them more details about the survey. In addition, they were sent travel diaries and a map of the United States to assist them in keeping track of key facts about their trips. It was hoped that this information would enhance the respondent's memory at the time of the interview.

1.1 Data Collection Schedule

Each sample household will be interviewed three or four times from April 1995 to February 1996 in order to collect each household's travel for all of 1995. Each of these interviews, called a cycle, will concentrate on trips taken during the time frame preceding the date of interview. That is, the first interview uses a reference period of January 1 to the date of interview and the second through fourth interviews use a reference period of the date of the previous interview to the date of the current interview. For example, if a household was interviewed on May 25, 1995 during the first cycle, the reference period for that interview was January 1, 1995 to May 25, 1995. If they were then interviewed on August 31, 1995 during the second cycle, the reference period would be May 25, 1995 to August 31, 1995.

Within each cycle the sample was divided into three representative samples called waves. A sample unit assigned to a wave stays in the same wave for the whole survey. Each wave is interviewed in a different month of the cycle to control the effect of recall on the sample results. This random assignment of the sample to specific waves throughout the course of the survey gives us the opportunity to evaluate the effect of recall on trip estimates. This will provide valuable information when designing future travel surveys. Section 2. includes a discussion of the work planned on the recall evaluation. Refer to Table 1 for the data collection schedule.

1.2 Pre-Data Collection Telephone Research

The retired CPS sample had telephone numbers for 79 percent of the sample addresses. To obtain the most up-to-date telephone numbers for the sample addresses, we sent the entire ATS sample to a private telephone research company called Telematch. Telematch found telephone numbers for 37 percent of the sample addresses. We also updated area codes where they had

changed since the time of the last CPS interview. At this point interviewers at the three telephone centers used CD-ROM phone disks to look for a telephone number for the sample address, a business with an address close to the sample address, an apartment rental office on the same street, or next door neighbors who could give us information about the sample address.

Table 1: ATS Data Collection Schedule

Cycle	Wave 1	Wave 2	Wave 3
1	4/26-5/31	6/1-6/30	7/1-7/31
2	8/1-8/31	9/1-9/30	10/1-10/31
3	11/1-11/30	12/1-12/31	1/1-1/31
4	1/15-2/15	2/1-2/29	No Data Collection

In January and February of 1995, interviewers at the telephone centers contacted the households by telephone and conducted the Post Mail Initial Call Screener (PMICS) operation. During this call, the interviewers verified the sample address and telephone number. They asked if the household had received the survey information (letter of introduction, travel diaries and a map). The interviewers also answered any questions the household member had regarding the survey and tried to obtain a contact person for the upcoming interview. During this call the interviewers encouraged the contact person and other household members to participate in the survey and to use the travel diary. PMICS also had a Telephone Research Operation (TRO) that tried to obtain telephone numbers for cases that had incorrect telephone numbers.

1.3 Centralized/Decentralized Interviewing Assignment

After the PMICS operation was complete, the sample was split into one of three strata based on the PMICS outcome. One stratum consisted of cases with confirmed addresses and phone numbers. The second stratum consisted of cases where one or both of these were not confirmed but there was reason to believe that they could be correct. For example, we reached an answering machine. The third stratum consisted of cases that we felt certain the phone numbers were incorrect. For example, we reached a recorded message

that said the number was no longer in service and we could not find a new number for that address.

At this point we determined that our sample of confirmed addresses and phone numbers (i.e., stratum one cases) was not sufficient. Without more sample cases in this stratum, we could not make full use of the centralized CATI facilities. We would then face budget problems because we would need to do more personal visit interviews than we originally planned. This was a costlier option which would have caused us to settle for a smaller sample size.

We then decided to add additional retired CPS sample cases which were last interviewed in January, February, July, August, September and October 1994. We refer to this as the new sample. Since this sample was added late, these cases did not go through the PMICS operation but, like the original sample, they did receive the introductory letters, a diary and a map. Instead of the PMICS outcome, we used the presence of a CPS telephone number to determine the stratum classification for the new sample. Since the new sample cases retired from CPS in 1994, we expect the phone numbers to be fairly accurate. Cases that had a CPS phone number were put into the first stratum. The remaining cases were put into the second stratum. Refer to Table 2 for the distribution of sample cases by stratum.

Table 2: ATS Sample Counts by Stratum

	Stratum 1	Stratum 2	Stratum 3
Old Sample (PMICS)	50,500	5,000	5,400
New Sample (PMICS)	17,700	1,400	

We then designated cases in each of these strata for centralized CATI or decentralized CATI/CAPI. Two-thirds of the cases in stratum one (i.e., cases with both the address and phone number confirmed) were sent to centralized CATI and the remainder were sent to decentralized CATI/CAPI. One fourth of the cases in each of the other two strata were sent to decentralized CATI/CAPI. The remaining units in these two strata were not sent for interview. We are increasing the weights on the decentralized CATI/CAPI cases in these

strata to represent the cases from the second and third strata that were not used.

1.4 Data Collection Procedures

The telephone centers attempted to complete CATI interviews with as many of their cases as possible. Typically, for Census Bureau surveys that conduct interviews from the telephone centers, cases that the CATI telephone centers cannot interview are sent out to the field to be interviewed by FRs (i.e., recycled). Since these cases weren't recycled for the ATS, these centralized CATI noninterviews will be represented by comparable cases in the decentralized CATI/CAPI workload. We will adjust the weight on the decentralized CATI/CAPI cases identified as matches to the centralized CATI noninterviews (i.e., "dual-frame" weighting).

To ensure comparability of centralized and decentralized components, we are using a maximum CATI approach from the FR's homes. We gave the FRs some guidelines to use so their workloads are completed using methods as similar as possible to the procedures used at the telephone centers.

Since this split-sample design with dual-frame weighting is a new approach for Census Bureau surveys, an evaluation of the effectiveness of this approach in conducting a survey is being designed. We will use the results of this evaluation for future survey planning. A discussion of the proposed evaluation is included in this paper in Section 3.

Beginning with the second interview, we will use a data collection technique called bounding. Use of this technique attempts to prevent trips reported in a previous interview from being reported again in a subsequent interview. When the respondent reports a trip, the interviewer will check the last ten trips from the prior interview to see if the trip was reported previously. If the new trip has the same destination and is in the same time frame as the previously reported trip, the interviewer will probe to see if they are reporting the same trip.

The ATS also utilizes a reinterview program that includes a data quality check. A quality control check is planned on a sample of the decentralized CATI/CAPI households to verify that the field representatives actually conducted interviews. In addition, a response-error check will be attempted on a sample of interviews to verify the quality of the data collected. The

reinterview program is discussed in greater detail in Section 4.

2. ATS RECALL BIAS/TELESCOPING EVALUATION

The ATS is using a 3-month recall period between interviews with an initial 4, 5, and 6 month recall for the first interview (See the ATS interviewing schedule in the introduction). There are two factors resulting from the ability of people to remember their trips which affect the quality of the data - one is recall bias and the other is telescoping.

Recall bias refers to the fact that people may forget about trips from the time they took them to the time they're actually interviewed. Obviously, the more recent the trip the more likely they'll be able to report it. Thus, the longer the recall the more likely we'll be to miss trips.

Telescoping refers to the fact that people tend to report trips more recently than when they were actually taken. For example, someone may report a trip in March that they actually took in February. There are two types of telescoping errors - internal (i.e., within the reference period) and external (i.e., between reference periods).

We're doing several things to minimize the effect of recall bias and telescoping. First, we sent the respondents travel diaries (calendars) for them to record their trips making their recall better. Next, we called them in the beginning of the year (this was the PMICS operation) to encourage them to use the calendar, to let them know we would contact them again to collect their data, and to let them know we only wanted trips since January 1 (to decrease the possibility of external telescoping from December into January).

Finally, we'll use bounding for the 2nd through 4th interviews. Bounding is explained in section 1.4.

We plan to do several things to adjust for extended recall in the first interview as well as measure the effect recall bias and telescoping may have on the data.

2.1 Recall Adjustment for 1st Quarter Interviews

Since data from the first quarter will have much longer recall than the other interviews, we plan to make an adjustment to reduce the effect of this larger recall bias on data quality. Following is a brief description of the problem and how we plan to address it.

We'll collect the data in three 'waves' where each wave will be a representative sample by itself. Each wave will consist of about one-third of the total sample. The first interview for each wave will take place in May (wave 1), June (wave 2), and July (wave 3) 1995. The data we collect will represent trips from January 1 to the date of interview. Thus, for the data collected in May, there will be 4-month recall for the January data, 3-month recall for the February data, 2-month recall for the March data, and so on. Likewise, for the data collected in June there will be 5-month recall for the January data, 4-month recall for the February data, and so on. For the data collected in July, there will be an extra month of recall.

The adjustment we plan to make will give higher weights to data collected in May, since it has the smallest recall, and lower weights to data collected in June and July. The adjustment will probably vary by month. That is, the factors for January will favor May's data more than June's and July's. For February and March, the factors will still favor May's data but to a lesser degree. This won't correct for all the bias. However, it should reduce the bias since it will give more weight to the more reliable data.

2.2 Effect of Recall and Telescoping on Trip Estimates

Clearly, recall bias and telescoping will have the biggest effect on the data in the first quarter (i.e., January, February, and March). These months have the longest recall therefore the respondents will probably forget to report some trips. January will also suffer from the internal telescoping of trips into later months but it probably won't have many trips externally telescoped from previous months. We're hoping the effect of external telescoping of trips from the prior calendar year will be minimal since people will be able to differentiate between trips taken before and after the new year. Since people tend to take fewer trips in the first quarter, recall bias and internal telescoping could further reduce the number of trips.

The second and third quarters (and October) will all have the same recall bias effect (i.e., between 1 and 3 months). The second quarter may have more internal telescoping than the third quarter but less than the first. Beginning with the second interview (i.e., data collected in August for the period May through the date of interview), we'll use bounding to diminish the external telescoping effect. But the data collected in June and July will still suffer from internal telescoping of trips from prior months into the later months.

Data for November and December will have less recall bias than other months. All data will be, at most, two months recall. We're hoping that respondents will be able to differentiate trips before or after the end of the year. However, they could report trips ending in 1996 in December and omit trips ending in December because they thought the trip ended in January. Thus it's not clear what the effect on the data will be.

2.3 Measuring the Effect of Recall and Telescoping on the Data

We plan to see how different assumptions about recall bias and telescoping affect trip estimates. We'll use three different sources to develop these assumptions. First, we'll use results from the response error reinterview. The response error reinterview is reconciling the return date of trips during the reinterview. We'll use the information from this operation to develop assumptions about internal telescoping. Second, we'll use data from November and December. These months will have less recall bias and less internal telescoping than other months. Finally, we'll use results from previous research on recall bias and telescoping for other surveys (e.g., the National Crime Victimization Survey and the National Survey of Fishing, Hunting, and Wildlife Associated Recreation).

To measure the effects of recall bias and telescoping, we plan to compare trip estimates from different waves for each month. For example, we'll compare trip estimates for the month of February using data collected in May, June, and July. This will represent a comparison of 3 vs. 4 vs. 5 month recall. We'll assume differences in these estimates that are larger than sampling error represent the net effect of recall bias and telescoping. Since both recall bias and telescoping affect monthly data, it will be difficult to separate the two effects. For example, in February the respondent may forget some trips he took but he may also report some trips from January in February.

Table 3 summarizes the comparisons we'll make.

We'll also compare trip estimates from January through April for households contacted during PMICS (old sample) with those that weren't contacted during PMICS (new sample) to see if the PMICS contact had any effect on recall bias and telescoping.

Limitations - In addition to recall bias and telescoping, there are four other effects which will limit the usefulness of our results. The first of these is bounding which was described at the beginning of this section.

Bounding won't eliminate external telescoping so the first month in each reference period may still be affected by this.

The second effect occurs because respondents will be asked about trips completed since the last interview. For example, the people interviewed in August will be asked about trips since their May interview instead of since May 1. Thus, data for some months (e.g., May for wave 1) will have two different lengths of recall.

The third effect occurs because new households who move in and for noninterviewed households we'll ask for trip data from the beginning of the previous quarter rather than the date of the last interview. For example, for new households interviewed in August we'll collect trip data for trips since April 1. Thus, data for some months may have two different recall lengths for this reason as well.

The fourth effect is that trips reported by new households will not benefit from bounding nor the possible use of a travel diary. Thus, they may have a larger recall bias as well as more external telescoping.

To see how these four factors affect the data, let's look at the monthly trip estimate from May. The May trip estimates will come from data collected in May, June, July, and August. Below is a description of the factors affecting data from these three data collection periods.

For wave 2 data collected in June, there is probably not much internal telescoping of trips from May into June. There will, however, be internal telescoping of trips from prior months into May. There is probably minimal loss of trips due to recall. New households that move in between the first and second interview will have May trips reported in September (i.e., longer recall than the rest of the wave 2 data).

For wave 3 data collected in July, there is internal telescoping of trips both into May from previous months and internal telescoping of trips from May to June and/or July. There is also more loss due to recall bias.

Wave 1 data were collected in both May and August with more than half of the data collected in August. For the data collected in May, there could be internal telescoping of trips from previous months into May. This is consistent with data from wave 2 and 3 described above. However, there won't be any external telescoping of trips from May into a later month. For data collected in August, the opposite will be true. There shouldn't be much external telescoping of trips

from previous months into May because we'll use bounding. However, there could be internal telescoping of trips from May into later months in the reference period. This is different from the data from waves 2 and 3. In addition, new households will have telescoping of trips both into and out of May since we'll get their trip data from April 1.

In summary, some of the comparisons, the recall bias, and the telescoping effect for each month, could vary from one wave to the next.

3. CATI/CAPI EVALUATION

As previously mentioned, we are using a split-sample design with "dual-frame" weighting for ATS. About 45,000 cases are assigned to centralized CATI interviewing in the Census Bureau's three telephone centers. About 35,000 cases are assigned to decentralized CATI/CAPI interviewing by the Census Bureau's field representatives (FRs).

The telephone center workloads only include cases with "good" telephone numbers as determined by two definitions or criteria of "good". The current telephone numbers were obtained for 3/4 of these cases during the PMICS operation that was implemented for the CPS sample that retired in 1991-93. The remaining cases were the new supplemental sample cases from the CPS sample that retired in 1994 which had CPS telephone numbers.

The cases that can't be contacted by the telephone centers will not be sent out to be interviewed by FRs. The weight for comparable cases identified in the decentralized CATI/CAPI workload will be increased to account for these noncontacted cases (i.e., dual-frame weighting).

The FR workloads will include a representative sample of all addresses. The sample will include (1) "good" telephone numbers, (2) addresses with telephones for which we couldn't get the current phone number or couldn't verify the number matched the correct address, and (3) addresses without telephones. This workload will include samples of both definitions of "good" telephone numbers.

We will implement a maximum CATI approach from the FR's homes. We will give the ATS FRs guidelines so their workloads are completed in a way that mirrors as closely as possible the methods used at the telephone centers. These guidelines will ensure the comparability

of the centralized and decentralized components which should facilitate accurate dual-frame weighting.

In addition, we will also collect some information/details on attempts to reach these decentralized cases by telephone as well as the reason a personal visit interview was conducted. This information is used to identify the comparable decentralized CATI/CAPI cases that would not have been reached by centralized CATI (i.e., "non-telephone" or "hard-to-reach-by-telephone" addresses that were originally "good" telephone numbers). The weight for these cases is increased to represent the centralized workload that couldn't be contacted.

3.1 Objectives of the Evaluation

ATS is the Census Bureau's first survey to use this unique approach of a split-sample design with "dual-frame" weighting. As a result, we feel it is critical to evaluate the different facets and implications of this approach. The results of this evaluation will be used to guide us in determining the optimal allocation of units with known telephone numbers to the centralized CATI and decentralized CATI/CAPI modes for the next ATS as well as for the Census Bureau's other demographic surveys.

Specifically, this evaluation will address the following issues:

- how well do the comparable decentralized cases we identified represent the noncontacted centralized cases?
- how effective is centralized CATI interviewing relative to the decentralized interviewing?
- how well do centralized CATI interviewed households represent the universe of households with "good" telephone numbers?
- how well do centralized CATI interviewed households represent the universe of all households with telephones?
- how well do centralized CATI interviewed households represent all households?

The above issues are addressed separately for both definitions of "good" telephone cases.

3.2 Dual-Frame Weighting Evaluation

In ATS dual-frame weighting, certain decentralized CATI/CAPI cases, that were originally classified as "good" telephone cases, were identified to represent noncontacted centralized CATI cases. These cases were identified by using the information we collected about the attempts to reach them by telephone as well as the reason a personal visit interview was needed. In essence, we attempted to identify decentralized CATI/CAPI cases that would not have been reached by centralized CATI. These selected decentralized CATI/CAPI cases had their weights increased to represent noncontacted centralized CATI cases.

Our major concern about this step is that the resulting sample may not accurately represent the universe. We are introducing bias from errors in determining which units in the decentralized CATI/CAPI sample would not have been reached by centralized CATI. In this evaluation, we plan to look for evidence related to the magnitude of "multiplicity bias" resulting from this step. This evaluation will consist of the following comparisons for the cases originally classified as "good" telephone cases:

- a comparison of original CPS demographic characteristics (e.g., family income, tenure, race, age and sex of householder, size of household) for noncontacted centralized CATI cases with those of decentralized CATI/CAPI cases used to represent them.
- a comparison of ATS demographic characteristics (e.g., family income, tenure, race, age and sex of householder, size of household) for contacted centralized CATI cases with those of decentralized CATI/CAPI cases identified as cases centralized CATI could have contacted.
- a comparison of major ATS travel characteristics (e.g., household trips and person-trips by mode of transportation, main reason for trip, length of trip, round-trip distance, and travel party size) for contacted centralized CATI cases with those of decentralized CATI/CAPI cases that centralized CATI could have contacted.

The above comparisons are made at the state, region, and US levels separately for each of the samples that used the two definitions of "good" telephone numbers.

3.3 Effectiveness of Centralized CATI

ATS was the Census Bureau's first survey where the telephone centers used a maximum effort to contact their entire workload of good phone cases without recycling cases to the decentralized staff. As a result, we want to evaluate the relative effectiveness of centralized CATI interviewing without recycling. This evaluation will consist of the following:

- a comparison of interview rates for centralized CATI sample to the % of telephone interviews for "good" telephone cases assigned to decentralized CATI/CAPI.
- a comparison of interview rates for centralized CATI sample to the % of telephone interviews for decentralized CATI/CAPI cases identified as those the telephone centers could have contacted.
- a comparison of the number of cases completed by each of the telephone centers as well as over all three telephone centers for each week of the interviewing period to see if a point exists at which interviewing is no longer effective.
- an analysis of major ATS travel and demographic characteristics of centralized CATI interviews that are completed after a large number of attempts to see if there is anything unique about these cases (e.g., a high or low incidence of travel).
- a comparison of centralized CATI cases costs (e.g., cost per case, cost per interview) with those for decentralized CATI/CAPI.

3.4 Representation of the Universe of Households with "Good" Telephone Numbers

The ATS split-sample design allows us to evaluate how well centralized CATI results represent the universe of households with "good" telephone numbers, using the two different definitions of "good". This evaluation will consist of the following:

- a comparison of ATS demographic characteristics (e.g., family income, tenure, race, age and sex of householder, size of household) for centralized CATI interviews with those of decentralized CATI/CAPI

interviews originally classified as "good" telephone cases.

- a comparison of major ATS travel characteristics (e.g., household trips and person-trips by mode of transportation, main reason for trips, length of trips, round-trip distance, and travel party size) for the centralized CATI interviews with those for decentralized CATI/CAPI interviews originally classified as "good" telephone cases.

The above comparisons are made at the state, region, and US levels separately for each definition of "good".

3.5 Representation of the Universe of All Households with Telephones

The ATS split-sample design allows us to evaluate how well centralized CATI results represent the universe of all households with telephones. This evaluation will consist of the following:

- a comparison of ATS demographic characteristics (e.g., family income, tenure, race, age and sex of householder, size of household) for centralized CATI interviews with those of all decentralized CATI/CAPI interviews with telephones.
- a comparison of major ATS travel characteristics (e.g., household trips and person-trips by mode of transportation, main reason for trips, length of trips, round-trip distance, and travel party size) for centralized CATI interviews with those for all decentralized CATI/CAPI interviews with telephones.

The above comparisons are made at the state, region, and US levels separately for centralized CATI interviews from each definition of "good".

3.6 Representation of the Universe of All Households

The ATS split-sample design allows us to evaluate how well centralized CATI results represent the universe of all households (i.e., households with telephones as well as households without telephones). This evaluation will consist of the following:

- a comparison of ATS demographic characteristics (e.g., family income, tenure,

race, age and sex of householder, size of household) for centralized CATI interviews with those of decentralized CATI/CAPI interviews.

- a comparison of major ATS travel characteristics (e.g., household trips and person-trips by mode of transportation, main reason for trips, length of trips, round-trip distance, and travel party size) for centralized CATI interviews with those for decentralized CATI/CAPI interviews.

The above comparisons are made at the state, region, and US levels separately for centralized CATI interviews from each definition of "good".

4. ATS RESPONSE ERROR AND QUALITY CONTROL REINTERVIEW

The ATS reinterview program consists of two components; Response Error (RE) reinterview and Quality Control (QC) reinterview. We selected a separate sample for each reinterview. This allowed us to tailor the sample design and methodology based on the goals of each reinterview. As a result, we hope to get better estimates of response error in addition to a quality control operation. This is an improvement on past reinterviews where response error and quality control were often combined into one reinterview. (i.e., one sample design and one methodology)

The ATS response error reinterview is set up to allow assessment of data quality. By measuring the accuracy of the number of trips reported, response variance for the number of trips reported and response variance for the estimates covering trip details, we can get a clearer picture of data accuracy and reliability.

The ATS quality control reinterview performs a quality control check on FRs to deter and detect falsification. While this part of reinterview is an integral part of Census Bureau demographic surveys, the sponsor is most interested in the accuracy of the number of trips reported. Therefore, this paper will focus on RE reinterview.

4.1 ATS Response Error (RE) Reinterview

The response error reinterview program began in August of 1995.

We want to measure:

1. accuracy in the number of trips reported,
2. reliability (response variance)
 - of number of trips reported, and
 - of the estimates covering trip details for trips reported in both the original interview and reinterview.

4.1.1 Methodology

We plan to conduct RE reinterviews of original CAPI cases by CATI. (Note: Any original CAPI cases without a phone number cannot be reinterviewed for RE.) All RE reinterviews are completed within two weeks of the original interview.

The reinterview respondent responds only for themselves. Proxy respondents are not acceptable because we primarily want to measure accuracy, and self-response is likely to yield responses closer to the truth.

At the beginning of RE reinterview, we ask the respondent if someone in the household completed portions of the travel diary. We also ask if anyone referred to the diary during the original interview. If the respondent indicates that the travel diary is complete, we prompt the respondent to get the diary for answering the reinterview questions.

Following the diary questions, we re-ask most of the original interview questions. Next, we reconcile any differences in the trips reported.

The CATI reinterview instrument automatically matches trips reported in the original interview with trips reported in the reinterview. Any unmatched trips are reconciled by the reinterviewer. The reinterviewer reconciles by probing about trips that were reported in the original interview or reinterview, but not both, and by making a determination based on the answers.

For ATS we plan to reconcile lists of trips (trips reported in the original vs. trips reported in reinterview) as opposed to reconcile close-ended questions. Although other organizations have undertaken automated reinterview with reconciliation, this is the Census Bureau's first attempt.[1]

As Blair and Sudman suggest, reinterview, and in this case reconciliation, may run the risk of offending the respondent as well as increasing the burden on the respondent.[2] At the end of the ATS survey and reinterview, we will assess the effect of respondent

burden by looking at ATS and reinterview response rates.

4.1.2 Sample Design

We selected the ATS RE sample from only original CAPI cases to avoid potential non-response bias in our response error estimates. CATI cases without a phone number or cases requiring a personal visit are not interviewed for the original survey (no recycles). CAPI cases without a phone number are interviewed by personal visit where the FR will try to obtain a phone number.

The ATS RE sample is essentially self-weighting, except for within household respondent selection. The respondent is a randomly selected adult household member (18 years or older) from the household roster, who responds for themselves. A self-weighting design indicates that each reinterview case has a constant baseweight. Use of a self-weighting design simplifies the calculation of response error estimates.

4.1.3 Advantages of Response Error Reinterview

Listed below are some of the advantages of ATS response error reinterview.

Self-Response

Since the reinterview respondent will respond only for themselves, we have a better chance of getting the truth when estimating bias.

Independent Response Variance Reinterview

The ATS response variance reinterview is independent, unlike many of the Census Bureau's paper and pencil reinterviews. In the past, using paper and pencil, the reinterviewer could flip back to the original responses before or during the reinterview. Because the ATS response variance reinterview is automated, the reinterviewer does not have access to previous responses when re-asking original survey questions. This allows us to fulfill the response variance reinterview model requirement of independence.[3] The reinterviewer has access to original data when reconciling trips.

4.1.4 Limitations

Listed below are limitations of ATS response error reinterview. These limitations are primarily due to operational and budget constraints.

Non-Response Bias

Since we are reinterviewing by CATI only (no recycles), non-response bias is possible. This bias arises because households without a phone or that require a personal visit are not reinterviewed.

Limited Number of Trips With Trip Details

Because respondents may report trips in a different order in reinterview, we'll collect trip details on all trips reported in reinterview. We collect trip details in the original interview for only the first ten trips reported. Therefore, we can measure response variance on the first ten trips reported in the original interview.

Reconciled Reinterview Not the Best Method to Measure Accuracy

A reconciled reinterview is not necessarily the best method to measure accuracy. It can be argued that comparing to administrative records or other comparable methods results is a better estimate of accuracy. However, for the ATS a reconciled reinterview seems most appropriate because no such travel records are available.

Any reinterview that replicates the original interview will not detect consistent response error. To detect consistent bias generating response error, we need better questions, better procedures or some other methodological improvement in reinterview compared to the original interview.

Reinterview Not an Exact Replication

The response variance reinterview model requires that reinterview be an independent replication of the original interview. The ATS response error reinterview does not entirely replicate the original interview. First, we are not interviewing by the same mode (CATI vs. CAPI). Secondly, the original interview may be done by proxy; whereas, the reinterview must be self-response. It is also worth noting that the respondent may remember their answers from the original interview and just repeat them in reinterview.

4.2 ATS Quality Control (QC) Reinterview

As mentioned earlier, a quality control check on FRs to deter and detect FRs who may be falsifying data is an integral part of the Census Bureau's demographic surveys. The quality control reinterview program began in June 1995. Since the primary purpose of the ATS

reinterview is to measure response error, we will briefly discuss the ATS QC reinterview program.

The QC reinterview sample is separate from the response error reinterview sample. Most ATS QC cases will be reinterviewed by phone, with personal visits allowed only when absolutely necessary. We first select a sample of FRs for QC reinterview. We then select cases from each selected FR's current assignment.

Among FRs who falsify, inexperienced FRs tend to falsify more frequently than experienced FRs. Therefore, we select inexperienced FRs at a higher rate than experienced FRs. However, more cases are selected for reinterview from an experienced FR's assignment.[4]

We conduct QC reinterview only on cases interviewed in the field. Telephone center interviews are often monitored making a quality QC reinterview unnecessary.

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.

Acknowledgements

The authors would like to thank Phil Gbur, Mark Hendrick, Tom Moore and Susan Dorinski for commenting on this paper.

References

- [1] Brick, M. et al. (1994). "A Study of Selected Nonsampling Errors in the 1991 Survey of 1989-1990 College Graduates." (Prepared under contract to the National Center for Education Statistics) Rockville, MD: Westat, Inc.
- [2] Blair, J. and Sudman, S. (1993). "Respondent Perceptions of Reinterviews." Presented at the 1993 Census Bureau Annual Research Conference.
- [3] Biemer, P. et al. (1991). "Measurement Errors in Surveys." John Wiley and Sons, Inc., New York.
- [4] Wetzel, A. (1993). "Falsification by Field Representatives." Internal Census Bureau Report. May 10, 1993

Table 3: American Travel Survey Recall Comparisons

Monthly Estimate for:	Month Interviewed (Number of Months Recall in Parenthesis)		
	Wave 1	Wave 2	Wave 3
January	May (4 months)	June (5 months)	July (6 months)
February	May (3 months)	June (4 months)	July (5 months)
March	May (2 months)	June (3 months)	July (4 months)
April	May (1 month)	June (2 months)	July (3 months)
May	August (3 months)	June (1 month)	July (2 months)
June	August (2 months)	September (3 months)	July (1 month)
July	August (1 month)	September (2 months)	October (3 months)
August	November (3 months)	September (1 month)	October (2 months)
September	November (2 months)	December (3 months)	October (1 month)
October	November (1 month)	December (2 months)	January (3 months)
November	January (2 months)	December (1 month)	January (2 months)
December	January (1 month)	February (2 months)	January (1 month)